Modeling Heterogeneous Peer Assortment Effects using Latent Class Pseudo-Maximum Likelihood Exponential Random Graph Models

Teague R. Henry

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology and Neuroscience.

Chapel Hill
2015

Approved by:

Kathleen Gates

Daniel Bauer

Peter Mucha

ABSTRACT

TEAGUE R. HENRY: Modeling Heterogeneous Peer Assortment Effects using Latent Class Pseudo-Maximum Likelihood Exponential Random Graph Models.
(Under the direction of Kathleen Gates)

This thesis develops a class of models for inference on networks called Sender/Receiver Latent Class Exponential Random Graph Models (SRLCERGMs). This class of models extends the existing Exponential Random Graph Modeling framework to allow analysts to model unobserved heterogeneity in the effects of nodal covariates and network features. Simulations across a variety of conditions are presented to evaluate the performance of this technique, and an empirical example regarding substance use among adolescents is also presented. Implications for the analysis of social networks in psychological science are discussed.

To my partner Jenn, whose support and dedication are my inspirations.

ACKNOWLEDGMENTS

Kathleen Gates, whose advising and assistance was invaluable. Daniel Bauer and Peter Mucha, whose comments made this thesis significantly stronger. My colleagues in the Gates Lab, who were always willing to chat. The Quantitative Psychology program at large, for teaching me and supporting the work on this thesis.

TABLE OF CONTENTS

LIST OF TABLES

1 INTRODUCTION

Network methods have emerged as an increasingly prominent tool in both psychological science and the broader social sciences. For example, recent research in adolescent development, behavior, and health utilize network methods to study the association between peers and a variety of behaviors such as adolescent alcohol use and smoking (Osgood, Ragan & Wallace, 2013; Mercken, Steglich, Sinclair & Holliday, 2012), childhood obesity (Shin, Valente, Riggs & Hu, 2014), deviant behavior (Osgood, Feinberg & Ragan, 2015), and sexual behavior (Ali & Dwyer, 2011). Any behavior or trait of interest to researchers that is associated with a social network structure is likely the result of two processes, influence processes, where the network structure (who is connected with whom) changes the behavior or trait of an individual over time and the reverse, selection processes, where the behavior or trait changes the network structure over time. These two processes are interrelated and together impact both network structure and the individual processes within the network such that, over time both processes can drive dynamic, multi-level shifts. Given this synergy between influence and selection processes, and their joint implications for change over time, network data analysis poses a variety of complex methodological problems.

However, in recent years, new methodologies have been developed to either simultaneously analyze peer selection and influence processes (e.g., SIENA; Snijders, de Bunt, & Steglich 2010), or focus on one, such as peer selection processes (e.g., STERGMs, Krivitsky & Handcock, 2014). Of additional interest is peer assortment, that is, the cross-sectional pattern of peer nominations which are conditioned upon individual traits and covariates. Although cross-sectional network analysis confounds the effects of peer influence and peer selection (Kandel, 1978), it provides greater insight into the nature of a given network structure at any time-point. Additionally, models that investigate cross-sectional peer assortment are similar to models that assess peer selection processes, specifically the exponential random graph modeling framework (ERGM; Frank & Strauss, 1986; Wasserman & Pattison, 1996; Moriss, Handcock & Hunter 2008), and its longitudinal extension, the separable temporal exponential random graph modeling framework (STERGM; Krivitsky & Handcock, 2014).

What underlies all of these network analytic methods is their ability to assess the effect of individual level traits and covariates, such as personality characteristics or observed behaviors, on the network structure. Furthermore, in directed networks (i.e. networks in which edges have a beginning and end point; e.g., peer friendship networks where individuals nominate peers as friends), researchers can assess sender and receiver effects of covariates. These effects assess the impact of a behavior, such as substance use, on the tendency for individuals to send or receive edges. Breaking apart covariate effects into sender and receiver components may prove useful when investigating the impact of individuals behaviors on network structure (i.e., selection processes).

Parallel to network methods recent flourishing in the psychological and social sciences, methods that seek to model and capture individual (or ideographic) differences are also resurging and may benefit from a network approach. Modeling ideographic differences can be done a variety of different ways, depending on the research question at hand and also the research design. For example, an ideographic analysis of relations of effects (Molenaar, 2004) allows researchers to examine a single individuals pattern of behavior, while other methods test the aggregate relations between outcomes as moderated by some individual trait or other behavior (Bauer, 2011). Of the existing individual differences methods, moderation by an observed variable is by far the most commonly used (Bauer, 2011). When modeling individual differences, methods such as multiple regression or ANOVAs are the typical statistical techniques used. However, current methods that analyze whole networks can also model observed variable moderators, typically through a multiplicative interaction effect. As such, further work that focuses on improving cross-sectional and selection network processes may be valuable for ideographic research, as these network methods would allow researchers to examine individual differences in social structure and effects.

This is even more vital, given that failure to model individual differences based on observed or unobserved variables can be significantly detrimental to researchers confidence in the validity of their data . For example, neglecting to include moderation effects in non-network data may result in biased estimates (Jaccard & Turrisi, 2003) and even spurious random effects (Bauer & Cai, 2009) It follows that in network analysis, failure to account for heterogeneity could have potentially severe consequences, especially because certain parameter combinations can cause the estimation to converge on a degenerate solution (Chatterjee & Diaconis, 2013). This risk of degeneracy is increased when estimating a mis-specified model, such as the failure to account for individual differences in effects.

Given burgeoning interests in networks and individual differences, a growing number of researchers are seeking to examine the role of moderators and individual differences in networks. For example, in adolescent peer influence work, researchers focus on peer selection and peer influence processes, and on how individual differences in traits and behaviors can shape network structure as a whole. Within this body of literature, the greatest amount of research has examined peer influence processes (for review see, Brechwald & Prinstein, 2011), with little work exploring potential moderators of peer selection or peer assortment (with exception; see Kiuru et al., 2010) Furthermore, there have been no studies that have used a whole network analysis method such as exponential random graph modeling or SIENA (Snijders, de Bunt, & Steglich, 2010) to study the moderation of peer influence, selection, or assortment by individual characteristics. However, this paucity of research is not due to a lack of research interest, but rather that current methodologies for testing network-based moderation effects in peer selection or peer influence do not allow for moderation in the estimates of behavioral or endogenous effects on network structure that is based on unobserved or latent variables.

Latent class models and mixture models are also commonly used in psychological research to assess individual differences based on unobserved data, and thus may provide a solution to this methodological hindrance. Specifically, finite mixture linear regression (DeSarbo & Cron, 1988) and finite mixture general linear regression (Wedel & DeSarbo, 1995) can be used to partition a sample into groups, in which each group has different relations among observed variables. This potentially models individual differences that are not due to an observed variable. One advantage to using a finite mixture regression approach in the presence of unobserved individual differences is that the regression models for each mixture component can be a combination of component-specific parameters as well as general parameters for which no individual differences are proposed.

In network modeling, latent variables and mixtures are commonly defined as latent communities, or groups of individual entities (be they humans, neurons, etc.) who tend to nominate each other more than they nominate entities in other groups. A great deal of methodology has been developed in the last decade to assess latent community structures, chief of which are heuristic algorithms that partition networks (for review, see Fortunato, 2010). There are several statistical methods for detecting latent communities, such as stochastic block modeling (Nowicki & Snijders, 2001), latent space modeling (Hoff, Raftery & Handcock, 2002), and more recently, the ERGM approach (Schweinberger & Hancock, 2014); however, all of these approaches focus on assessing latent community structure and do not focus on assessing individual differences in the effects of covariates based on unobserved subgroups.

By contrast, the present thesis aims to extend the definition of latent classes and mixtures on networks, and model latent classes of individuals who differ in their effect of observed variables on their position in the network. To accomplish this, this thesis develops a sender/receiver finite mixture modeling approach for ERGMs, focusing on cross-sectional or peer assortment models. This work will help address the methodological conundrum facing researchers who wish to utilize a whole network analysis method to test the moderation of peer influence, selection, or assortment by individual differences. ERGMs are a class of exponential family inferential models that allow the modeling of the effect of covariates on network structure, as well as modeling the complex dependencies in networks. The form of an ERGM makes it amenable to a mixture modeling approach. A set of simulations will be presented that demonstrate the sender/receiver finite mixture ERGMs ability to recover the mixture components, as well as the consequences of failing to model the individual differences. Finally, an empirical example examining individual differences in the effect of alcohol use on the network structure in a middle-school will be presented.

1.1 Network Terminology and Model Description

In order to work with networks as a data structure, the terminology used to describe networks in this thesis needs to be established.

A network consists of a set of *nodes* and a set of *edges* that connect the nodes together. Nodes define the members of a network, and can represent anything that has relations with other things. Nodes commonly represent individuals, corporations, townships, network routers, and brain regions of interest (ROIs). Edges define the relationship between nodes in the network, such as friendships, board interlock, information exchange, or functional connections. Edges can be *directed* (e.g., friendship nominations among adolescents) or *undirected* (e.g., board interlock among companies). To store edges, and to represent them in a fashion that amenable to analysis, an *adjacency matrix* is used. An adjacency matrix, A, is a $N \times N$ matrix where the A_{ij} is the edge value between each unique pair of individuals

i and j for all N individuals.

This data structure allows the analyst to represent any network structure, be it directed or undirected, weighted or un-weighted in a portable fashion. A directed edge A_{ij} is the relationship from node i to node j, and does not inform on the relationship from node j to node i. An example of an directed edge would be a friendship nomination. The nomination of friendship from individual i to individual j does not necessitate the nomination of friendship from individual i to individual j. Other networks have undirected edges, where $A_{ij} = A_{ji}$. One common example of this is the correlation network between regions of interest (ROI) in the brain. A correlation is a bi-directional effect, and the corresponding edge between ROI i and ROI j will have the same value from regardless of the direction of the edge.

Friendship nominations and brain ROI correlations also provide examples as to differing types of edges. The edges corresponding to friendship nominations take on the value of 0, for no friendship nomination, or 1 for a friendship nomination. This type of edge is referred to as *binary*. Binary networks reflect the presence or absence of a connection between nodes. The brain ROI correlations take on any value between -1 and 1. These connections are referred to as *weighted* edges. The edges in a weighted network can take on any real number value.

In addition to the presence or absence of edges, and possible weights on edges, often times *node-level attributes* are also included in the analysis. Node level attributes are simply values that reference particular nodes. Gender is a node level value for a adolescent friendship network, while size of an ROI is a node level variable for a brain ROI correlation network. In this manuscript, the $N \times P$ matrix of node level variables will be denoted using X.

The present thesis focuses on directed, binary graphs that take into account node-level attributes.

1.2 A Brief Description of Exponential Random Graph Models

Exponential Random Graph Models (ERGMs; Wasserman & Pattison, 1996) are a family of models used to analyze network structure both contingent on endogenous predictors as well as on covariates of the individual nodes. ERGMs model the entire network as a sample from a population of networks with a set of sufficient statistics associated with the population. These sufficient statistics can include the number of edges in the network, the number of edges between individuals with dissimilar values on one specific covariate and/or the number of closed triangles in the network, to name a few examples. These models were originally developed to analyze *dyad independent* networks, where the probability of an edge between two nodes was only dependent on the characteristics of those two nodes (Holland & Leinhardt, 1981). They were extended by Frank and Strauss (1986) to include transitivity effects, where the probability of an edge existing depends on the existence of neighboring edges.

The modeling of transitivity proves to be difficult, and only in recent years has there been estimation procedures developed to successfully model transitivity (Hunter & Handcock, 2006). Since the development of a efficient estimation procedure, ERGMs have been actively researched in several fields, and have been extended in a variety of ways. A longitudinal extension was recently developed (STERGMs; Krivitsky & Handcock, 2014) that model the formation and dissolution of ties over multiple observations of a network. Additionally, there has been work on modeling latent community structures within ERGMs. Schweinberger and Handcock (2015) developed a Bayesian procedure for analyzing very large networks, something that ERGMs typically struggle with. They succeed by modeling a large network as a mixture of smaller sub-networks.

Increasingly in the past decade, methodology to model heterogeneity and individual differences in networks has been under development. Heterogeneity in networks is commonly described as a *latent community* structure. Broadly speaking, latent community structure is a type of individual difference as different nodes (individuals) are following different rules regarding association. However, it is an overly restrictive form of heterogeneity, as it doesn't allow for nodes to have behave differently *and* not group into community structures. Instead, a broader definition of a heterogeneity in network structure will be used in this manuscript, that is a network is heterogeneous if the nodes that make up the network behave differentially.

There are two dominant statistical methods for analyzing latent classes as latent communities in random networks. Those are the Latent Space Model (Hoff, Raftery & Handcock, 2002) and the Stochastic Block Modeling framework of (Nowicki & Snijders, 2001). The latent space model posits that every node is located in a latent social space, and the nodes connectivity to other nodes is dependent on the location of each node in the social space. This method can be used to find community structure and implicitly accounts for transitivity and the effects of covariates, however does not allow any heterogeneity in the effects of the covariates on network structure.

The stochastic block modeling framework posits that a network structure is entirely due to the latent community structure, and if the latent community structure is known, then the edges within the network are dyad independent given community assignment. Furthermore, this framework models the network as a mixture of Erdős-Rényi graphs (Erdős & Rényi, 1961) (networks where each edge is distributed i.i.d Bernoulli), which is a even more restrictive assumption than dyadic independence. This framework has been extended in recent years most notably with the work of Daudin (2008), where the authors develop a mixture model on random graphs that allow for community structure to be defined in several different ways.

In this thesis, we use the exponential random graph modeling framework to construct latent class models. Exponential random graph models are a model based inferential system for analyzing networks, one of the two that are in use now. The other method, stochastic actor based modeling (Snijders, de Bunt & Steglich, 2010) is primarily used for longitudinal data, and relies on agent based simulation for estimation. Due to these properties, the focus of development will be on exponential random graph models.

As a model based inferential system, ERGMs require the analyst to specify which effects they want to model. This provides fine grained control to the analyst, however, the price is that these effects are homogeneous across all nodes in the network, in that every node given the nodal covariates behaves the same. It is fairly apparent that real-data networks are not completely homogeneous, given the recent interest in latent community detection and other methods that explicitly assess for heterogeneity within a network. This interest suggests that a model that allows for heterogeneities and individual differences in ERG modeling would be of great assistance. Although it is fairly simple to model moderation between observed variables using multiplicative interactions in ERG modeling, there has been no methodology to date that allows for the modeling of latent category moderation, in which the nodes in a network are partitioned into a set of latent classes, and each class potentially has a different effect of nodal covariates and network features. This modeling of unobserved heterogeneity is similar to the approach taken in finite mixture regression (Wedel & de Sarbo, 1995)

The model that is developed in this thesis is termed a Sender/Receiver Latent Class ERGM. This class of models posits that every node has a latent class assignment, and this latent class assignment moderates the effect of specific nodal covariates on the probability of edges directed towards the node (A Receiver Latent Class ERGM), or edges directed away from the node (A Sender Latent Class ERGM), but not both. This division often aligns with hypotheses made by researchers who are interested in associations between nodal attributes and nominations. The choice to restrict the model to either assess the effect of either sender latent class or receiver latent class was made to render the model tractable to estimation via a hard type EM algorithm (Kearns, Mansour & Ng, 1998; MacQueen, 1967; Dempster, Rubin & Laird, 1977) for reasons elaborated below.

In what follows, ERGMs and the Sender/Receiver Latent Class ERGMs will be explained in technical detail.

1.3 Exponential Random Graph Models

An exponential random graph model models the observed network as a sample from a exponential family distribution with the following form. Let $\mathbf{A}$ be a random directed network of size N with sample space $\mathcal{A}$, and $\mathbf{X}$ a $N \times P$ matrix of fixed and known nodal covariates, with $\mathbf{a}$ being a sample realization of $\mathbf{A}$. Then the distribution of $\mathbf{A}$ is

$$P(\mathbf{A} = \mathbf{a}|\mathbf{X}) = \frac{\exp[\langle \boldsymbol{\theta}, s(\mathbf{a}, \mathbf{X})\rangle]}{\psi(\boldsymbol{\theta})} \tag{1.1}$$

Where $\boldsymbol{\theta}$ is a d length vector of natural parameters in log-odds metric, $s(\mathbf{a}, \mathbf{X})$ is a d length vector of sufficient statistics, $\langle \boldsymbol{\theta}, s(\mathbf{a}, \mathbf{X})\rangle$ is the inner product of the vector of natural parameters and the vector of sufficient statistics and $\psi(\boldsymbol{\theta})$ is a normalizing constant. The log-likelihood of the parameters (Hunter & Handcock, 2006) is then:

$$\ell(\theta) = \langle \boldsymbol{\theta}, s(\mathbf{a}, \mathbf{X})\rangle - \log(\psi(\boldsymbol{\theta})) \tag{1.2}$$

As was outlined above, the vector of sufficient statistics $s(\mathbf{a}, \mathbf{X})$ can contain a variety of terms. Sufficient statistics that will be used in the simulation studies and the empirical examples are described below. Broadly, these can be divided into two categories, sufficient statistics that are wholly based on the network structure, structural statistics, and statistics that are based on the node level attributes, node level attribute statistics.

1.3.1 Structural Statistics

What follows are the technical definitions as defined in the statistical literature.

Edges. The parameter estimate associated with the edges statistic in any ERGM acts as an intercept for the model. The sufficient statistic described by this term is a count of the number of edges in the network.

Mutuality. This sufficient statistic is used when modeling directed networks, and the sufficient statistic is the total number of reciprocated dyads (Holland & Leinhart, 1981). These are dyads where $a_{ij} = a_{ji} = 1$. The parameter associated with this statistic can be described as the tendency for edges to be reciprocated.

Geometrically Weighted Edgewise Shared Partners (GWESP). This sufficient statistic is used to account for transitivity in the network (Hunter, Goodreau & Handcock, 2008). Transitivity is the phenomena where an edge is more likely to be present if the edge has shared partners. This is commonly referred to colloquially as "A friend of a friend is my friend." Transitivity is a common phenomena in real-world networks, and failure to model transitivity can result in bias in the estimates of other effects (Van Duijn, Gile, & Handcock, 2009). However, when a term that models transitivity is added, this induces a dependency structure in the network that prevents a direct maximization of the likelihood during estimation. This dependency structure also runs the risk of causing degeneracy in estimation, where the expected value of the networks distribution converges to a completely empty or completely full network. Details of why this occurs can be found in Chatterjee and Diaconis (2014). The geometrically weighted edgewise shared partners term works to prevent degeneracy by down-weighting large numbers of shared partners. A GWESP term has both an effect parameter $\theta^{(GWESP)}$ and a weight parameter τ. τ ranges from 0 to ∞, with 0 meaning that the effect of shared partners does not increase beyond having 1 shared partner, and ∞ meaning that the effect of additional shared partners is linear with no asymptote.

If τ is allowed to be freely estimated, then the ERGM becomes a member of a curved exponential family (Efron, 1978), which complicates estimation. In the simulations that follow, τ is considered fixed and known, a practice that is done in substantive analysis where the curved exponential family models fail to estimate. (Hunter, Goodreau & Handcock, 2008)

1.3.2 Nodal Attribute Terms

All terms below are described in Pattison and Robins (2001).

Attribute Matching. The parameter estimate associated with this sufficient statistic is used to evaluate the effect of nodes matching on a categorical variable such as gender or ethnicity. The sufficient statistic is the number of edges between nodes that match on the categorical variable.

Attribute Outdegree The parameter estimate associated with this sufficient statistic is used to evaluate the effect of a continuous variable on the number of edges coming out of nodes. The node level sufficient statistic is the outdegree of a node multiplied by the attribute value for that node.

Attribute Indegree Similarly to the Attribute Outdegree term, the parameter estimate associated with this sufficient statistic reflects the effect of a continuous variable on the number of edges nodes in the network receive. The node level sufficient statistic is indegree of the node multiplied by the attribute value for that node.

Attribute Absolute Difference The parameter associated with this sufficient statistic evaluates the effect of similarity on a continuous variable between nodes on the probability of an edge. The edge level sufficient statistic is the absolute difference in the values of the attribute for the sender and receiver of the edge multiplied by the presence or absence of that edge.

1.4 Estimation for ERGMs

Estimation of the parameter vector $\boldsymbol{\theta}$ for an ERGM model is complicated when a term that measures transitivity is included in the model. ERGMs, as originally proposed by Holland and Leinhardt (1981), did not contain transitivity terms, and instead were considered *dyad independent* models. As was outlined above, *dyad independent* models are ones where the state of a dyad (what edges in a dyad are active), are conditionally independent of all other dyads given the characteristics of the nodes that make up the dyad in question. This simplifies the likelihood of a *dyad independent* ERGM to that of a multinomial regression for a directed network (the categories being the 4 possible states of a directed dyad), or a binary logistic regression for an undirected network, which in turn makes estimation of a *dyad independent* ERGM trivial.

However, when a term that accounts for transitivity, such as a GWESP term, is added to the model the resulting dependency structure changes the likelihood into something that is less tractable to estimation. The core of the problem with transitivity assessing ERGMs was first realized by Frank and Strauss (1986) and stems from the definition of the normalizing constant $\psi(\boldsymbol{\theta})$. In a general ERGM, $\psi(\boldsymbol{\theta})$ is defined as such:

$$\psi(\boldsymbol{\theta}) = \sum_{a \in \mathcal{A}} \exp[\langle \boldsymbol{\theta}, s(\mathbf{a}, \mathbf{X}) \rangle] \tag{1.3}$$

Where $\mathcal{A}$ is the sample space of the distribution of the random graph A. This sum over all networks of a given size is intractable for any reasonable sized network. For example, the number of possible directed networks of 25 nodes is 2^{25^2-25} [1] or approximately $4.4 * 10^{180}$. Iterating over this number of networks in a reasonable time frame is impossible for any computing system currently in existence. An approximate solution to this estimation problem was proposed by Hunter and Handcock (2006), which uses Markov Chain Monte Carlo sampling to obtain a sample of networks at a given set of parameters, and then maximize the likelihood of the parameter values using that sample to approximate the normalizing constant. This procedure, when applied iteratively, converges to the maximum likelihood estimates of the parameters $\boldsymbol{\theta}$ as the number of iterations approach ∞. There

1 For a directed network of size N, each node has a possible number of edges equal to N-1, as self edges are not allowed. As such, the maximum number of edges in a directed network of size N is $N * (N - 1)$ or $N^2 - N$. If the network is binary, then each edge can have a value of 0 or 1. As such, using permutation with replacement, the number of possible networks (each of which can be considered a $N^2 - N$ tuple) is 2^{N^2-N}

are several problems with this sampler, most notable of which is the difficulty in getting a sample of networks that properly covers the density of the distribution for approximating the normalizing constant. Estimation for ERGMs is an ongoing field of research.

Due to the iterative nature of the approximate ML estimator of Hunter and Handcock (2006), it is less than ideal to implement in an Expectation Maximization framework, which in of itself is iterative. Additionally, the MCMC sampler does not scale well to larger networks, limiting the practical size of a network to less than 300 nodes. An alternative to the ML estimator appears in the maximum pseudolikelihood approach of Frank and Strauss (1986). Very recently there was a deterministic approximation to the normalizing constant that operates in quadratic time, which is promising for estimating ERGMs on larger networks (Pu et al., 2015).

1.4.1 Maximum Pseudolikelihood Estimation of ERGMs

When transitivity terms were first introduced by Frank and Strauss (1986), the estimator proposed for ERGMs was a pseudo likelihood approach first inspired by Besag (1974) pseudolikelihood approach for spatial models.

Recall that the distribution of observed variable ERGM is:

$$P(\mathbf{A} = \mathbf{a}|\mathbf{X}) = \frac{\exp[\langle \boldsymbol{\theta}, s(\mathbf{a}, \mathbf{X})\rangle]}{\psi(\boldsymbol{\theta})} \tag{1.1}$$

Consider now the conditional probability of an edge A_{ij} given the rest of the network, $\mathbf{A}_{ij}^c$:

$$P(A_{ij} = 1|\mathbf{A}_{ij}^c = \mathbf{a}_{ij}^c) = \frac{\exp(\langle \boldsymbol{\theta}, \mathbf{s}_{ij}\rangle)}{1 + \exp(\langle \boldsymbol{\theta}, \mathbf{s}_{ij}\rangle)} \tag{1.4}$$

Where $\mathbf{s}_{ij}$ is the change in the sufficient statistics when A_{ij} goes from 0 to 1. This expression is identical to a logistic regression model treating the $\mathbf{s}_{ij}$ as fixed and known covariates. The *pseudo-likelihood* approximation to the log-likelihood as developed by Frank and Strauss (1986) for an ERGM is then the likelihood for a binary logistic regression of the following form:

$$P(A = a|\boldsymbol{\theta}) \approx \prod_{(i,j)\in\mathbf{a}} \frac{(\exp(\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle))^{a_{ij}}}{1 + \exp(\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle)} \tag{1.5}$$

Taking the log of the above transforms it to the pseudo-log-likelihood which can be simplified as shown:

$$\begin{aligned}
\hat{\ell}(\boldsymbol{\theta}) &= \log\left[\prod_{(i,j)\in\mathbf{a}} \frac{(\exp(\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle))^{a_{ij}}}{1 + \exp(\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle)}\right] \\
&= \log\left[\frac{\exp(\sum_{(i,j):a_{ij}=1}\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle)}{\prod_{(i,j)\in\mathbf{a}} 1 + \exp(\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle)}\right] \\
&= \sum_{(i,j):a_{ij}=1} \langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle - \sum_{(i,j)\in\mathbf{a}} \log(1 + \exp(\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle))
\end{aligned}$$

Note that $\langle\boldsymbol{\theta}, s(\mathbf{a}, \mathbf{X})\rangle$ is equal to $\sum^{(k)} \theta^{(k)} s^{(k)}(\mathbf{a}, \mathbf{X})$, and that $s^{(k)}(\mathbf{a}, \mathbf{X})$ is equal to $\sum_{(i,j):a_{ij}=1} s_{ij}^{(k)}$. Therefore $\sum_{(i,j):a_{ij}=1}\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle = \langle\boldsymbol{\theta}, s(\mathbf{a}, \mathbf{X})\rangle$. This leads to the final expression of the pseudo-log-likelihood below:

$$\hat{\ell}(\boldsymbol{\theta}) = \sum_{(i,j)\in\mathbf{a}} \log(P(A_{ij} = 1|\mathbf{A}_{ij}^c)) = \langle\boldsymbol{\theta}, s(\mathbf{a}, \mathbf{X})\rangle - \sum_{(i,j)\in\mathbf{a}} \log(1 + \exp(\langle\boldsymbol{\theta}, \mathbf{s}_{ij}\rangle)) \tag{1.6}$$

Note that this pseudo-log-likelihood differs from the log-likelihood only in the normalizing constant. This pseudo-log-likelihood is identical to that of a binary logistic regression, and can be maximized by established means.

There are several issues with using the MPLE for ERGMs that model transitivity. Lubbers and Snijders (2007) provide evidence that the standard errors of the estimates in an MPL fit ERGM tend to be underestimated, and the whole model tends to overfit to the data. Van Duijn, Giles and Handcock (2009) show in a simulation study that the estimates of the transitivity parameter from an MPL fit ERGM are substantially less efficient than that of a ML fit ERGM. However, recently it has been proposed that the properties of the MPL estimator have not been thoroughly studied (Chatterjee & Diaconis, 2013), and that there is evidence for its asymptotic normality (Comets & Janzura, 1998).

Although it is entirely possible to use MLE when obtaining latent class solutions, there are several issues that suggest MPLE is a better choice for estimation. As will be shown, the latent class solution for a sender/receiver latent class ERGM is estimated using the EM algorithm. Using the MLE during the maximization step is computationally intensive, with running times an order of magnitude above the running time of MPLE. Furthermore, in light of the possibility of degeneration, repeated applications of the MLE while using non-optimal latent class solutions increases the probability that at least one iteration will be degenerate, which would cause estimation to fail. As such, in current thesis, we use MPLE during the maximization step of an EM algorithm implementation. Given the known issues with MPLE, we use MLE to obtain a final set of parameter estimates, once the EM algorithm has converged.

1.5 Sender/Receiver Latent Class ERGMs

1.5.1 Latent Class Generalized Linear Models

In 1988, Wedel and Desarbo developed a mixture model for generalized linear models. This approach relies on the hard EM type algorithm for estimation, otherwise known as the k-means algorithm (MacQueen 1967, Dempter, Rubin & Laird, 1977), and provides an elegant solution to the estimation problem in this thesis. As the MPL for an exponential random graph model is identical to the likelihood for a binary logistic regression, Wedel and Desarbo's approach can be applied.

The joint log pseudolikelihood for a Sender Latent Class ERGM as derived for the first time in this thesis is:

$$\hat{\ell}(\mathbf{Z},\boldsymbol{\theta}) = \sum_i \sum_{q=1}^{Q} Z_{iq}\log(\alpha_q) + \sum_i \sum_{j\neq i} \left[a_{ij} \left[\sum_{q=1}^{Q} Z_{iq}\langle \boldsymbol{\theta}_q, s_{ij}\rangle \right] - \sum_{q=1}^{Q} Z_{iq}\log(1+\exp(\langle \boldsymbol{\theta}_q, \mathbf{s}_{ij}\rangle)) \right] \tag{1.7}$$

Where Z_{iq} is 1 if node i is in class q and 0 otherwise. α_q is the marginal probability of a node being in class q. $\boldsymbol{\theta}_q$ is a vector of parameters for class q. a_{ij} is the value (0 or 1 for a binary network) of the edge from node i to node j. s_{ij} is the vector of changes in the sufficient statistics of the network if edge a_{ij} went from 0 to 1, given the rest of the network. Any parameter θ may be held equal across all classes, and thus rendered homogeneous.

1.5.2 Expectation Step

The general expectation for a latent class general linear model (where the probability densities are exponential family) is in the following form (Wedel & DeSarbo, 1995, Eq 7):

$$E[z_{iq}|\boldsymbol{\theta},\mathbf{y}_i] = \frac{\alpha_q \prod_{k=1}^{K} f_{ik|q}(\mathbf{y}_{ik}|\theta_{\mathbf{q}})}{\sum_{q=1}^{Q} \alpha_q \prod_{k=1}^{K} f_{ik|q}(\mathbf{y}_{ik}|\theta_{\mathbf{q}})} \tag{1.8}$$

Where $\mathbf{y}_i$ is a vector of observations (each of which can come from a different arbitrary probability density) for case i and $f_{ik|q}(.)$ is the probability density function for the kth element of the ith case given the ith case is in class q.

Here we can express the expectation for the sender latent class exponential random graph model in precisely that form of Equation 1.8.

$$E[z_{iq}|\boldsymbol{\theta},\mathbf{a}] = \frac{\alpha_q \exp(\sum_{j\neq i} [a_{ij}\, [\langle \boldsymbol{\theta}_q, s_{ij}\rangle] - \log(1+\exp(\langle \boldsymbol{\theta}_q, s_{ij}\rangle))])}{\sum_{q=1}^{Q} \alpha_q \exp(\sum_{j\neq i} [a_{ij}\, [\langle \boldsymbol{\theta}_q, s_{ij}\rangle] - \log(1+\exp(\langle \boldsymbol{\theta}_q, s_{ij}\rangle))])} \tag{1.9}$$

Note that in the above likelihood, edges that share the same sender node all are assigned the same latent class. Intuitively, this is due to the assignment of the latent class being sender based rather than edge based. This maps onto the Wedel and DeSarbo approach, as they classify observation vectors. In this case the observation vector is the vector of all possible edges a node could and did send.

With the expectation derived, we diverge from the approach laid out in Wedel and DeSarbo (1995). In Wedel and DeSarbo's approach, the expectations produce probabilities of class membership, and these probabilities are then used in the maximization step. This approach is known as mixture EM, which is the EM algorithm laid out originally by Dempster, Rubin and Laird (1977). In this thesis, we take a classification likelihood approach. The classification likelihood approach (for examples see: Symons, 1981; McLachlan & Peel, 2005), assigns each case the class label that has the highest probability. This hard assignment of classes implies that although the likelihood of the data will be a mixture across the classes, each case is not a mixture across the latent classes. Classification likelihood can be fit using an EM-type algorithm that instead of using the expected value of class membership (a mixture approach), instead uses the class label with the highest probability.

This hard assignment has advantages and disadvantages. Classification likelihood algorithms consistently find more informative latent classes than mixture algorithms in the sense that the KL divergence of the mixture components is maximized, however does not maximize the likelihood of the data as the mixture EM algorithm does (Kearns, Mansour & Ng, 1998). Additionally, classification likelihood works well for small sample sizes (Celeux & Govaert, 1993). There are several disadvantages to classification likelihood. Classification likelihood does not recover ill-separated mixture components well, nor does it recover very unbalanced group sizes (Celeux & Govaert, 1993; Govaert & Nadif, 1996). This failure to recover ill-separated groups directly comes from the maximization of the KL divergence. If the groups are ill separated, classification likelihood tends to model the data using a single distribution in which the majority of cases are placed. This can be an advantage if the presence of latent classes is being tested, however, if the presence of latent classes is assumed, this approach will not detect them if they are ill separated. Finally, classification likelihood does not exhibit optimal large sample properties in that it tends to be asympototically biased (Bryant, 1991), though this bias appears to be lessened when allowing for different cluster sizes (Celeux & Govaert, 1993). In this thesis, classification likelihood is used for both its computational benefits (in that the likelihood easily separates on a case by case basis), as well as its property of maximum informativeness as measured by KL divergence. Finally, the classification likelihood approach used in this thesis does not assume equal cluster sizes. This choice tends to lessen the bias inherit in using classification likelihood. In larger samples however, mixture maximum likelihood estimation should be used to take advantage of the asymptotic consistency inherit in that approach.

1.5.3 Maximization Step

With the latent class labels assigned, the conditional likelihood is:

$$\hat{\ell}(\boldsymbol{\theta}|\mathbf{Z}, \mathbf{a}) = \sum_{(i,j)\in\mathbf{a}} \left[a_{ij} \left[\sum_q Z_{iq} \langle \boldsymbol{\theta}_q, \mathbf{s}_{ij} \rangle \right] - \sum_q Z_{iq} \log(1 + \exp(\langle \boldsymbol{\theta}_q, \mathbf{s}_{ij} \rangle)) \right] \tag{1.10}$$

Where Z_{iq} is a binary indicator of node is membership in latent class q. This likelihood is the same as the likelihood for a binary logistic regression, and as such can be maximized in θ using standard ML estimation, specifically iteratively weighted least squares.

1.5.4 A Note on Standard Errors

In Wedel and DeSarbo's (1995) description of a mixture model on the general linear model, they note that the standard errors of the estimates can be calculated using a weighted information matrix, taking into account the uncertainty of latent class assignments. However, this relies on the mixture EM approach. In a classification EM-like approach, the parameters are estimated at each maximization step with each node assigned to a class with a weight of 1. As such, the standard errors of the estimates can be interpreted as the standard error of the estimated parameters if the estimated latent class solution was observed and true. Again, estimation with a mixture EM approach is entirely tenable for this model, in which case the standard errors would be weighted.

1.5.5 Estimation for Sender/Receiver Latent Class ERGMs

There are two additional considerations for estimating Sender/Receiver Latent Class ERGMs. The first is that of multiple start values. The EM algorithm, either classification or mixture type is susceptible to local maxima (Hipp and Bauer, 2006; Rubin, Dempster and Laird, 1977). To account for this, we initialize the estimation with multiple start values. Once every model has converged, the latent class solution with the greatest likelihood is selected.

Additionally, there are the multiple issues with using the MPLE to obtain parameter estimates. To account for this, once the EM algorithm has converged the latent class labels are used to estimate an MLE ERGM. The parameter estimates of that model are presented.

The full estimation algorithm proceeds as follows:

1. Initialize starting values for latent class labels $\mathbf{Z}_0$

2. For iteration k

   (a) Maximize the likelihood $\ell(\theta_k|Z_{k-1})$ to obtain θ_k

   (b) Compare $\ell(\theta_k|Z_{k-1})$ against $\ell(\theta_k - 1|Z_{k-2})$. If change in log-likelihood is less than tolerance, estimation has converged to a solution.

   (c) Obtain the hard latent class labels using $E(Z_k|\theta_k)$

3. Once estimation has converged, save solution, initialize a new set of starting values and return to step 2.

4. Once solutions are saved for a pre-defined number of starting values, choose the converged solution with the greatest final $\ell(\theta|Z)$. This solution has the most probable set of latent class labels out of all of the solutions.

5. With the most probable solution selected, fit a MLE ERGM considering the latent class labels as observed, and return the parameter estimates θ.

2 SIMULATIONS

2.1 Methods

In order to test the performance of the Sender/Receiver Latent Class ERGM and to assess the consequences of neglecting to model latent class heterogeneities, a set of simulations were performed.

In order to more accurately reflect the type of data a researcher would encounter as well as to follow good practice in ERGM methodology development (van Duijn, Gile, & Handcock, 2009), the simulations were based off of an empirical dataset, consisting of a single network of 151 middle schoolers. In addition to directed friendship nomination, information on gender, ethnicity, tobacco use, alcohol use, marijuana use and antisocial behavior were collected. Heterogeneity on sender alcohol use, sender absolute difference in alcohol use, and the edge parameter was modeled in the empirical analysis. For detailed descriptive statistics, and results of the empirical analysis see Section 3.

Due to missing data at the covariate levels, multiple imputation was used to generate a set of covariate datasets for the empirical analysis in Section 3. In these simulations, a single dataset of covariates was randomly selected from the multiply imputed datasets, and was used for all simulated networks. This follows the simulation procedure outlined in Van Duijn, Giles and Handcock (2009). For each simulation trial, a new network was simulated without changing the covariate dataset.

There were 9 simulation conditions in total (See Table 2.1). Simulation conditions were arrived at by examining the parameter estimates from the empirical model and changing some of the parameter values to reflect different situations. For example, the empirical model indicated heterogeneity in the effect of both sender alcohol use and the edge term.

For condition 1 of the simulations, which only modeled a heterogeneous effect of the edge term, the parameter for sender alcohol use was set to be equal across both classes. Additionally, for all conditions the edge parameter was increased from the empirical example to simulate less sparse matrices and therefore increase the amount of information in the data. Finally, in all conditions, approximately 25% of nodes were in class 2. This was derived from the empirical results.

Four out of the 9 simulation conditions were replications of conditions, but with increased effect sizes. These increased effect size conditions are denoted with a +. Increased effect sizes were assessed to get a sense of what approximate effect sizes return accurate latent class labeling. Condition 1 and Condition 1+ simulated networks with two latent classes that differed only in the edge parameter. Condition 2 and Condition 2+ simulated networks with two latent classes that differed only in the effect of a sender nodal attribute (alcohol use) . Condition 3 and Condition 3+ simulated networks with two latent classes that differed only in the effect of sender nodal attribute and the edge parameter. Condition 4 and Condition 4+ simulated networks with two latent classes that differed only in the GWESP effect. Note that the homogeneous edge effect for Conditions 4 and 4+ was increased from -3.85 to -3.65 to ensure that the total edge count in Condition 4 and 4+ was comparable to the edge count in other conditions. Finally, Condition 5 generated networks with no latent class structure. Simulation parameters are presented in Table 2.1.

In Conditions 1 through 4 and 1a though 4a, correctly specified models were fit to the data as well as homogeneous models. In Condition 5 a model that specified 2 latent classes and heterogeneity on the sender alcohol use parameter and the edges parameter was estimated, along with a homogeneous model.

Raw and relative bias were assessed, as well as computed standard errors from the simulation set and average estimated standard errors. Additionally, Rand Indices and Adjusted Rand Indices (Steinley, 2004) were computed for the comparison of the estimated latent

Table 2.1: Simulation Conditions: Values are the data generating parameters per condition.

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1) Class 1	0.97	0.97	0.97	0.97	0.97	0.97	1	1.2	0.97
GWESP (.1) Class 2	0.97	0.97	0.97	0.97	0.97	0.97	0.2	0.2	0.97
Mutual	2.34	2.34	2.34	2.34	2.34	2.34	2.34	2.34	2.34
Gender Match	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52
Ethnicity Match	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
Sender Tobacco Use	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Receiver Tobacco Use	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11
Sender Anti Social	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64
Receiver Anti Social	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
Sender MJ	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
Receiver MJ	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
Receiver Alcohol Use	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Abs Diff Tobacco Use	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21
Abs Diff MJ	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39
Abs Diff Anti Social	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08
Abs Diff Alcohol Use	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Sender Alcohol Use Class 1	0.06	0.06	0	0	0	0	0.06	0.06	0.06
Sender Alcohol Use Class 2	0.06	0.06	-0.25	-1	-0.25	-1	0.06	0.06	0.06
Edges Class 1	-3.5	-3.25	-3.85	-3.85	-3.5	-3.25	-3.65	-3.65	-3.85
Edges Class 2	-4.25	-4.5	-3.85	-3.85	-4.25	-4.75	-3.65	-3.65	-3.85

MJ: Marijuana, Abs Diff: Absolute Difference

class labels and the true latent class labels. This indicates performance of the method, in that if the method failed to detect true latent class labels for a condition, the method would have not performed well with that parameter set. Both Rand Index and Adjusted Rand Index are presented. Rand Index is a raw measure of agreement between two sets of labels, while the Adjusted Rand Index is a measure of agreement that accounts for the expected level of agreement if the labels were assigned randomly. Finally, Rand and Adjusted Rand are presented for the latent class labels of only the nodes that had a value higher than 0 for alcohol use. These results are presented as there are several conditions where the latent classes are in part defined on the effect of sender alcohol use. If alcohol use is 0 for an individual, and the latent classes are defined wholly on alcohol use, then the node with no alcohol use would be assigned to any latent class at random. By examining the subset of nodes with positive alcohol use, we can examine the performance of the method when the latent classes are defined only for a subset of nodes.

All conditions had 500 networks simulated. The networks were simulated using the MCMC approach used in the R package statnet (Handcock, Hunter, Butts, Goodreau & Morris, 2003). This approach randomly toggles edges within a network with the probability according to the generating model. This allows the simulated data to properly reflect transitivity, as well as contain sampling variability. To clarify, only the networks themselves were simulated. The covariate data set was the same across all simulations as was the true latent class labeling.

Estimation of these models used twenty random start values per trial as per the algorithm described above.

2.2 Simulation Results for Heterogeneous Models

Table 2.2 contains the estimated parameters for the heterogeneous models, while Table 2.3 shows the mean Rand Indices and Adjusted Rand Indices for latent class recovery for every individual, as well as individuals who use alcohol. Table 2.4 contains the relative bias and Table 2.5 contains the raw bias. Table 2.6 presents average estimated standard error and standard deviation of the estimates across the simulated networks. Finally, Table 2.7 contains the mean probability of class membership, and standard deviation of class membership for correctly and incorrectly classified nodes across all conditions. As described below, these simulations indicate that the model can recover true latent class structure, and gives an indication as to what level of class difference would be necessary.

Table 2.2: Mean Estimated Parameters calculated across 500 trials in each condition

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1) Class 1	0.997	0.979	0.982	0.994	0.999	0.980	0.892	1.091	1.017
GWESP (.1) Class 2	-	-	-	-	-	-	0.221	0.246	-
Mutual	2.340	2.340	2.340	2.338	2.339	2.337	2.340	2.339	2.342
Gender Match	0.519	0.523	0.518	0.521	0.519	0.521	0.513	0.511	0.522
Ethnicity Match	0.837	0.848	0.844	0.838	0.836	0.852	0.821	0.819	0.854
Sender Tobacco Use	0.173	0.174	0.128	0.119	0.153	0.134	0.155	0.161	0.125
Receiver Tobacco Use	-0.114	-0.121	-0.124	-0.132	-0.126	-0.131	-0.117	-0.117	-0.128
Sender Anti Social	-0.458	-0.594	-0.709	-0.785	-0.511	-0.620	-0.466	-0.524	-0.594
Receiver Anti Social	0.226	0.227	0.242	0.252	0.242	0.228	0.232	0.226	0.237
Sender MJ	0.271	0.308	0.300	0.336	0.265	0.307	0.254	0.259	0.335
Receiver MJ	0.208	0.212	0.216	0.212	0.201	0.196	0.212	0.209	0.212
Receiver Alcohol Use	0.043	0.040	0.034	0.029	0.040	0.048	0.039	0.036	0.039
Abs Diff Tobacco Use	-0.207	-0.202	-0.196	-0.189	-0.195	-0.191	-0.199	-0.199	-0.194
Abs Diff MJ	-0.392	-0.391	-0.397	-0.390	-0.385	-0.380	-0.391	-0.387	-0.391
Abs Diff Anti Social	-0.046	-0.067	-0.087	-0.121	-0.052	-0.081	-0.040	-0.041	-0.077
Abs Diff Alcohol Use	-0.013	-0.011	-0.004	-0.001	-0.009	-0.016	-0.012	-0.010	-0.011
Sender Alcohol Use Class 1	0.077	0.064	0.005	-0.001	0.024	0.000	0.090	0.083	0.053
Sender Alcohol Use Class 2	-	-	-0.262	-0.987	-0.209	-0.997	-	-	0.046
Edges Class 1	-3.691	-3.299	-3.845	-3.836	-3.672	-3.277	-3.721	-3.684	-3.773
Edges Class 2	-4.384	-4.528	-	-	-4.368	-4.519	-	-	-4.097

MJ: Marijuana, Abs Diff: Absolute Difference

Table 2.3: Mean Rand Indices and Adjusted Rand Indices (Standard Deviations computed from the sample) computed across 500 trials within each condition. Results indicate best recovery occurs in the + conditions, with higher class separation. Structural effects such as GWESP are also recovered well.

Condition	RI	ARI	RI Alcohol Users	ARI Alcohol Users
1	0.773 (.064)	0.541 (.126)	0.766 (.089)	0.532 (.179)
1+	0.963 (.025)	0.924 (.051)	0.945 (.052)	0.89 (.104)
2	0.573 (.055)	0.06 (.120)	0.636 (.056)	0.272 (.137)
2+	0.591 (.086)	0.1 (.178)	0.851 (.054)	0.703 (.108)
3	0.817 (.060)	0.629 (.120)	0.899 (.069)	0.798 (.137)
3+	0.98 (.017)	0.959 (.034)	0.998 (.007)	0.997 (.014)
4	0.841 (.048)	0.667 (.096)	0.839 (.076)	0.678 (.153)
4+	0.932 (.034)	0.862 (.069)	0.924 (.056)	0.848 (.113)

Rand Indices and Adjusted Rand Indices were computed as the empirical mean and empirical standard deviation over every trial in a given condition. These are not analytic results from a hyper-geometric distribution and are not meant to test for significance. For class recovery, the mean adjusted Rand indices contained in Table 2.3 indicate that across all simulation conditions, increasing the difference between the latent classes leads to increased recovery of the latent classes. However, for conditions with smaller class separation (Conditions 1, 2, 3, 4), the recovery of the latent classes was not optimal. This is likely due to the use of the classification likelihood, which doesn't perform well for classes that are ill-separated. For most conditions, the greatest improvement in recovery of the latent class labels occurred only when looking at alcohol users. Conditions 2 and 2+ reveals even at larger differences between the latent classes, recovery of the class labels for alcohol users is still not as good as for other conditions (ARI of .272 and .703 for Conditions 2 and 2+

respectively). This is likely due to the latent class being defined solely on individuals who used alcohol, a subset of the sample, as opposed to other conditions that had the latent classes also be defined by structural parameters such as edges or GWESP.

These results indicates caution when using a latent class ERGM, as the effect size for a latent class that is wholly defined by a covariate needs to be larger than the effect size for a latent class that is defined by covariates and network structures. Additionally, with the use of classification likelihood, caution should be taken when analyzing a network to ensure that theoretically the classes are well separated.

One set of results of note is that Condition 4 and 4+ had remarkably good recovery of the latent class labels, even at at the lower class difference (ARI of .667 and .862 for Condition 4 and 4+ respectively). This suggests that heterogeneities in the GWESP term are reflected strongly in observed network structure. Finally, standard deviations of the Adjusted Rand Indices suggest that there was variability in recovery rates due to sampling fluctuations. This variability can lead, particularly in conditions with low class separation, to very poor rates of classification. As one would expect, there is more variability in the ARIs the lower the average ARI becomes, again, suggesting that the classification likelihood version of this method should be used when class separation is high.

Table 2.4: Relative Bias in the Parameter Estimates For Heterogeneous Models, calculated using Mean Raw Bias over True Parameter Value. Results indicate modest levels of bias in homogeneous parameters. This bias is in line with expected bias from the approximate MLE estimation

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1) Class 1	0.027	0.009	0.013	0.025	0.030	0.010	-0.108	-0.091	0.048
GWESP (.1) Class 2	-	-	-	-	-	-	0.107	0.228	-
Mutual	0.000	0.000	0.000	-0.001	-0.001	-0.001	0.000	0.000	0.001
Gender Match	-0.002	0.006	-0.003	0.002	-0.003	0.001	-0.013	-0.016	0.003
Ethnicity Match	-0.004	0.010	0.005	-0.002	-0.005	0.014	-0.023	-0.025	0.017
Sender Tobacco Use	0.332	0.340	-0.013	-0.088	0.175	0.032	0.191	0.240	-0.035
Receiver Tobacco Use	0.032	0.099	0.131	0.198	0.145	0.188	0.063	0.060	0.167
Sender Anti Social	-0.285	-0.071	0.108	0.227	-0.201	-0.031	-0.272	-0.182	-0.072
Receiver Anti Social	-0.059	-0.052	0.007	0.049	0.007	-0.051	-0.034	-0.058	-0.012
Sender MJ	-0.180	-0.067	-0.090	0.017	-0.196	-0.071	-0.230	-0.216	0.016
Receiver MJ	-0.010	0.010	0.029	0.010	-0.043	-0.065	0.007	-0.003	0.009
Receiver Alcohol Use	0.083	-0.011	-0.145	-0.263	-0.012	0.198	-0.036	-0.100	-0.013
Abs Diff Tobacco Use	-0.013	-0.039	-0.066	-0.100	-0.073	-0.093	-0.051	-0.054	-0.074
Abs Diff MJ	0.004	0.002	0.019	-0.001	-0.014	-0.026	0.002	-0.009	0.003
Abs Diff Anti Social	-0.425	-0.165	0.085	0.517	-0.353	0.011	-0.495	-0.486	-0.032
Abs Diff Alcohol Use	0.341	0.137	-0.570	-0.950	-0.055	0.638	0.232	0.036	0.141
Sender Alcohol Use Class 1	0.291	0.075	NA	NA	NA	NA	0.502	0.388	-0.110
Sender Alcohol Use Class 2	-	-	0.050	-0.013	-0.163	-0.003	-	-	-0.237
Edges Class 1	0.055	0.015	-0.001	-0.004	0.049	0.008	0.020	0.009	-0.020
Edges Class 2	0.031	0.006	-	-	0.028	-0.049	-	-	0.064

MJ: Marijuana, Abs Diff: Absolute Difference, NAs due to true effect being 0.

Red highlights indicates relative bias in homogeneous parameters above .1, blue highlights indicate relative bias in heterogeneous parameters above .1.

In terms of bias in the parameter estimates, there was good recovery of heterogeneous parameters. In Condition 4 and 4+, the relative bias for the recovery of the lower GWESP term (GWESP Class 2), which had a true value of .2 in both conditions, was the worst. The relative bias here was .107 for Condition 4 and .228 for Condition 4+. However, the raw bias (contained below in Table 2.5) in terms of magnitude was quite small, .021 and .046 respectively. This level of bias is not relevant to researchers analyzing empirical data, and is likely due to the GWESP term being difficult to estimate in general. The one other heterogeneous term that had a relative bias of above .1 was that of Sender Alcohol Use in Class 2 for Condition 3. The relative bias for this parameter was -.163, with a raw bias of .041. The true parameter value was -.25, which suggests that while the relative bias was high, the actual level of raw bias was comparatively low.

The homogeneous parameters recovery (such as gender match, ethnicity match, etc.) was reasonable with very few parameters having greater than .1 absolute raw bias or greater than .1 relative bias. However the Sender Anti Social effect for conditions 1, 2+, 3, 4, and 4+ all have raw bias magnitude greater than .1. This effect was held homogeneous across all conditions and had the highest magnitude of the homogeneous effects (-.64). The patterning of bias across all conditions was not consistent, and this likely indicates that the effect was being disrupted by the effects of the structural parameters such as GWESP. Additionally, with Sender Tobacco Use, several conditions had relative bias greater than .1. However, the true effect of Sender Tobacco Use was small (.13) in all conditions, which can lead to small levels of raw bias translating into large relative bias.

This occurrence of high relative biases for very small true parameters is expected and most notably occurs for the Absolute Difference of Anti-Social Behavior, Sender Alcohol use (for models with homogeneous Sender Alcohol Use), and Absolute Difference in Alcohol Use. Furthermore, for conditions 2, 2+, 3 and 3+ the true effect of Sender Alcohol Use in Class 1 was 0, which would lead to undefined relative bias for any level of raw bias.

As for Condition 5, recovery of homogeneous parameters was quite good, with no raw bias being greater than .1, and very few relative biases greater than .1. This suggests that the mis-specification of the latent class structure did not lead to systematic bias in the homogeneous parameters. Additionally, the spread of the latent class parameters around the true homogeneous parameters was quite reasonable, with Sender Alcohol Use having a true effect of .06, and the latent classes returning an effect of .053 and .046, while the Edge effect is 3.85 while the latent classes returned effects of 3.773 and 4.097 respectively.

Table 2.5: Mean Raw Bias in the Parameter Estimates for Heterogeneous Models, calculated across 500 trials in each condition

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1) Class 1	0.027	0.009	0.012	0.024	0.029	0.010	-0.108	-0.109	0.047
GWESP (.1) Class 2	-	-	-	-	-	-	0.021	0.046	-
Mutual	0.000	0.000	0.000	-0.002	-0.001	-0.003	0.000	-0.001	0.002
Gender Match	-0.001	0.003	-0.002	0.001	-0.001	0.001	-0.007	-0.009	0.002
Ethnicity Match	-0.003	0.008	0.004	-0.002	-0.004	0.012	-0.019	-0.021	0.014
Sender Tobacco Use	0.043	0.044	-0.002	-0.011	0.023	0.004	0.025	0.031	-0.005
Receiver Tobacco Use	-0.004	-0.011	-0.014	-0.022	-0.016	-0.021	-0.007	-0.007	-0.018
Sender Anti Social	0.182	0.046	-0.069	-0.145	0.129	0.020	0.174	0.116	0.046
Receiver Anti Social	-0.014	-0.013	0.002	0.012	0.002	-0.012	-0.008	-0.014	-0.003
Sender MJ	-0.059	-0.022	-0.030	0.006	-0.065	-0.023	-0.076	-0.071	0.005
Receiver MJ	-0.002	0.002	0.006	0.002	-0.009	-0.014	0.002	-0.001	0.002
Receiver Alcohol Use	0.003	0.000	-0.006	-0.011	0.000	0.008	-0.001	-0.004	-0.001
Abs Diff Tobacco Use	0.003	0.008	0.014	0.021	0.015	0.019	0.011	0.011	0.016
Abs Diff MJ	-0.002	-0.001	-0.007	0.000	0.005	0.010	-0.001	0.003	-0.001
Abs Diff Anti Social	0.034	0.013	-0.007	-0.041	0.028	-0.001	0.040	0.039	0.003
Abs Diff Alcohol Use	-0.003	-0.001	0.006	0.009	0.001	-0.006	-0.002	0.000	-0.001
Sender Alcohol Use Class 1	0.017	0.004	0.005	-0.001	0.024	0.000	0.030	0.023	-0.007
Sender Alcohol Use Class 2	-	-	-0.012	0.013	0.041	0.003	-	-	-0.014
Edges Class 1	-0.191	-0.049	0.005	0.014	-0.172	-0.027	-0.071	-0.034	0.077
Edges Class 2	-0.134	-0.028	-	-	-0.118	0.231	-	-	-0.247

MJ: Marijuana, Abs Diff: Absolute Difference

In summary, Tables 2.4 and 2.5 suggest that the sender latent class models fit to heterogeneous data do a reasonable job at recovering the model parameters. Bias in the parameter estimates is in part due to ERG modeling being approximate even when using the MCMC-MLE. A degree of bias is to be expected. That being said, the fact that bias still remained even when fitting the generating model to data suggests that researchers should have stringent criteria for interpreting results from latent class ERGMs, interpreting both effect size as well as significance level.

Table 2.6: Mean Estimated Standard Error calculated across 500 trials per condition | Simulation Error calculated from the difference between the parameter estimates and the true parameter estimates across 500 trials per condition - Heterogeneous Networks. Results indicate disagreement between estimated standard error and simulation error due to the implementation of classification likelihood estimation.

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1) Class 1	.153|.152	.140|.126	.172|.158	.148|.148	.137|.137	.122|.118	.086|.342	.099|.430	.207|.199
GWESP (.1) Class 2	-|-	-|-	-|-	-|-	-|-	-|-	.086|.348	.099|.434	-|-
Mutual	.070|.063	.072|.065	.068|.060	.071|.067	.072|.071	.077|.071	.096|.069	.094|.067	.068|.060
Gender Match	.039|.033	.040|.036	.037|.032	.038|.035	.039|.036	.042|.038	.056|.040	.056|.039	.037|.032
Ethnicity Match	.039|.038	.041|.038	.038|.033	.039|.035	.041|.038	.043|.040	.056|.044	.055|.041	.038|.034
Sender Tobacco Use	.075|.076	.074|.086	.087|.083	.092|.091	.084|.090	.081|.083	.107|.086	.103|.088	.082|.080
Receiver Tobacco Use	.073|.066	.073|.065	.086|.081	.091|.088	.082|.080	.079|.076	.106|.077	.101|.072	.081|.073
Sender Anti Social	.130|.198	.131|.150	.134|.129	.135|.136	.134|.200	.138|.145	.171|.200	.168|.161	.129|.150
Receiver Anti Social	.128|.113	.130|.119	.127|.117	.131|.128	.131|.127	.137|.127	.172|.132	.171|.120	.126|.110
Sender MJ	.074|.077	.077|.078	.076|.079	.097|.094	.082|.095	.107|.104	.108|.099	.109|.100	.076|.069
Receiver MJ	.074|.064	.077|.070	.076|.069	.098|.097	.083|.077	.108|.104	.108|.077	.109|.078	.075|.065
Receiver Alcohol Use	.042|.037	.043|.038	.045|.042	.056|.052	.046|.044	.053|.049	.058|.044	.057|.040	.042|.037
Abs Diff Tobacco Use	.073|.067	.072|.064	.086|.082	.091|.089	.082|.079	.079|.075	.106|.076	.101|.075	.081|.073
Abs Diff MJ	.068|.058	.071|.065	.071|.063	.095|.091	.078|.071	.105|.097	.104|.072	.106|.072	.069|.058
Abs Diff Anti Social	.113|.114	.115|.109	.111|.107	.115|.107	.116|.114	.122|.111	.162|.127	.160|.119	.113|.101
Abs Diff Alcohol Use	.035|.031	.036|.032	.039|.034	.051|.049	.040|.035	.047|.043	.051|.037	.051|.035	.035|.031
Sender Alcohol Use Class 1	.042|.055	.043|.051	.054|.145	.068|.498	.055|.136	.079|.505	.057|.059	.057|.056	.050|.049
Sender Alcohol Use Class 2	-|-	-|-	.053|.144	.068|.499	.055|.129	.076|.504	-|-	-|-	.051|.050
Edges Class 1	.189|.420	.174|.653	.207|.193	.183|.183	.173|.409	.155|.632	.132|.121	.143|.115	.246|.289
Edges Class 2	.189|.390	.174|.618	-|-	-|-	.173|.379	.156|.644	-|-	-|-	.246|.289

MJ: Marijuana, Abs Diff: Absolute Difference

Red Highlights indicate substantial disagreement between mean estimated standard error and simulation error.

As for estimated standard errors of the estimates and the standard deviations across the simulation trials for every homogeneous parameter the estimated standard error and the standard deviation of the parameter across the condition agreed. However, for the heterogeneous parameters the estimated standard error is considerably smaller than the simulation standard deviation (Table 2.6 highlighted to show). This is in part due to error in classification across the simulation trials. If there are simulation trials where the individuals are mis-classified then the estimates of the heterogeneous effects will be more variable than if in every trial the latent class labels were recovered with perfect fidelity. Additionally, the estimated standard errors assume that the latent class assignment is the true assignment, and are not adjusted for uncertainty in class assignment. This is a reflection of the hard-type EM algorithm in use in the estimation.

Table 2.7: Mean Posterior Class Probabilities For True Class Assignment for both correctly classified and incorrectly classified nodes (Standard Deviation), calculated over 500 trials for each condition. Results indicate that when nodes are correctly classified, they have a high probability of being in the correct class. When nodes are incorrectly classified, the probability of being in the correct class is substantially lower than .5.

Condition	Correctly Classified		Incorrectly Classified	
	$1 \rightarrow 1$	$2 \rightarrow 2$	$1 \rightarrow 2$	$2 \rightarrow 1$
1	0.958 (0.085)	0.928 (0.106)	0.197 (0.145)	0.225 (0.151)
1+	0.996 (0.027)	0.992 (0.036)	0.163 (0.154)	0.176 (0.154)
2*	0.949 (0.1)	0.929 (0.115)	0.134 (0.146)	0.149 (0.15)
2+*	0.997 (0.021)	0.988 (0.056)	0.08 (0.129)	0.067 (0.094)
3	0.972 (0.071)	0.94 (0.1)	0.217 (0.14)	0.238 (0.148)
3+	0.997 (0.023)	0.994 (0.032)	0.163 (0.139)	0.212 (0.154)
4	0.984 (0.052)	0.953 (0.096)	0.174 (0.158)	0.202 (0.142)
4+	0.994 (0.031)	0.981 (0.061)	0.181 (0.157)	0.191 (0.138)
5	0.769 (0.138)	0.758 (0.139)	0.244 (0.137)	0.232 (0.139)

*: Mean and Standard deviation based on nodes with Alcohol Use greater than 0.

The posterior class probabilities contained in Table 2.7 indicate that for nodes that were correctly classified, the posterior class probabilities for that classification were close to 1. This in turn suggests that for the most part, the latent class model correctly parsed the differences between the classes. However, for individuals that were mis-classified, the probability that individuals would be in their correct class was quite low across all conditions (approximately .2). Ideally, mis-classified nodes would have correct class probabilities close to .5, which would suggest issues with sampling variability. The standard deviations on the mis-classified probabilities are also quite high, suggesting that the correct classification probability ranged quite widely between trials. The high probability of correct classification as well as the small standard deviations lend support to this method doing a reasonable job in recovering the latent classes of the individuals.

Additionally we can infer from Table 2.7 that as a result of the use of classification likelihood, the latent classes were inferred to be quite separated from each other. This in turn results in the extreme pattern of mis-classification found in the low posterior probabilities of assignment to the correct class for nodes who where miss-classified.

Finally, it is important to note that the posterior class probabilities presented in Table 2.7 do not reflect the amount of mis-classification. For example, by the ARI Condition 2+ performed significantly better than condition 2, yet in Table 2.7 it appears that for individuals who were mis-classified they had lower probability of correct classification in condition 2+ than in condition 2. This is in part due to condition 2+ having better classification than condition 2, which forces those nodes which are mis-classified to have worse posterior class probabilities. This effect can be seen in each of the condition pairs.

The posterior class probabilities for condition 5 simply reflect that for that condition, nodes were classified into each group nearly randomly, so the errors in classification reflect the sizes of the latent classes.

2.3 Results for Homogeneous Models

What follows are the results from fitting homogeneous models to the data generated from the conditions. Mean parameter values are presented in Table 2.8. Relative biases are presented in Table 2.9 and Raw biases in the parameter estimates are presented in Table 2.10. Mean estimated standard error and standard deviations across the simulation trials are presented in Table 2.11. Note that for Conditions 1 through 4+, the homogeneous model fitted is not the generative model, while for Condition 5, the homogeneous model is the true model.

Table 2.8: Mean Parameters for Homogeneous Models calculated across 500 trials for each condition

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1)	1.097	1.290	0.993	1.069	1.142	1.622	0.535	0.586	1.006
Mutual	2.348	2.359	2.339	2.331	2.348	2.332	2.344	2.343	2.344
Gender Match	0.505	0.480	0.517	0.511	0.499	0.440	0.482	0.474	0.519
Ethnicity Match	0.785	0.710	0.844	0.806	0.767	0.619	0.716	0.692	0.852
Sender Tobacco Use	0.181	0.209	0.157	0.199	0.209	0.271	0.180	0.193	0.117
Receiver Tobacco Use	-0.097	-0.087	-0.111	-0.100	-0.090	-0.062	-0.094	-0.088	-0.128
Sender Anti Social	-0.347	-0.139	-0.539	-0.408	-0.274	0.001	-0.200	-0.113	-0.632
Receiver Anti Social	0.214	0.198	0.230	0.209	0.219	0.152	0.207	0.195	0.236
Sender MJ	0.234	0.157	0.299	0.302	0.193	0.143	0.188	0.155	0.327
Receiver MJ	0.203	0.201	0.214	0.208	0.198	0.201	0.209	0.207	0.209
Receiver Alcohol Use	0.036	0.022	0.030	-0.011	0.026	0.014	0.025	0.019	0.039
Abs Diff Tobacco Use	-0.218	-0.214	-0.208	-0.212	-0.219	-0.215	-0.200	-0.202	-0.194
Abs Diff MJ	-0.383	-0.365	-0.395	-0.389	-0.378	-0.348	-0.372	-0.365	-0.388
Abs Diff Anti Social	0.042	0.111	-0.029	0.073	0.089	0.215	0.085	0.113	-0.078
Abs Diff Alcohol Use	-0.011	-0.008	-0.002	0.041	0.003	0.024	-0.011	-0.009	-0.011
Sender Alcohol Use	0.109	0.140	-0.146	-0.467	-0.068	-0.164	0.119	0.135	0.059
Edges	-4.267	-4.618	-3.913	-4.061	-4.354	-5.088	-3.852	-3.843	-3.895

MJ: Marijuana, Abs Diff: Absolute Difference

Table 2.9: Relative Bias for Homogeneous Models calculated with Mean Raw Bias (Table 2.10) over True Parameter Values. Results indicate that there is a substantial and systematic pattern of bias in the homogeneous parameters when heterogeneity is not properly modeled.

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1) Class 1	0.131	0.330	0.024	0.102	0.177	0.672	-0.465	-0.512	0.037
GWESP (.1) Class 2	-	-	-	-	-	-	1.675	1.930	-
Mutual	0.003	0.008	0.000	-0.004	0.003	-0.003	0.002	0.001	0.002
Gender Match	-0.029	-0.077	-0.006	-0.017	-0.040	-0.154	-0.073	-0.088	-0.002
Ethnicity Match	-0.065	-0.155	0.005	-0.040	-0.087	-0.263	-0.148	-0.176	0.014
Sender Tobacco Use	0.392	0.608	0.208	0.531	0.608	1.085	0.385	0.485	-0.100
Receiver Tobacco Use	-0.118	-0.209	0.009	-0.091	-0.182	-0.436	-0.145	-0.200	0.164
Sender Anti Social	-0.458	-0.783	-0.158	-0.363	-0.572	-1.002	-0.688	-0.823	-0.013
Receiver Anti Social	-0.108	-0.175	-0.042	-0.129	-0.088	-0.367	-0.138	-0.188	-0.017
Sender MJ	-0.291	-0.524	-0.094	-0.085	-0.415	-0.567	-0.430	-0.530	-0.009
Receiver MJ	-0.033	-0.043	0.019	-0.010	-0.057	-0.043	-0.005	-0.014	-0.005
Receiver Alcohol Use	-0.100	-0.450	-0.250	-1.275	-0.350	-0.650	-0.375	-0.525	-0.025
Abs Diff Tobacco Use	0.038	0.019	-0.010	0.010	0.043	0.024	-0.048	-0.038	-0.076
Abs Diff MJ	-0.018	-0.064	0.013	-0.003	-0.031	-0.108	-0.046	-0.064	-0.005
Abs Diff Anti Social	-1.525	-2.388	-0.638	-1.913	-2.113	-3.688	-2.063	-2.413	-0.025
Abs Diff Alcohol Use	0.100	-0.200	-0.800	-5.100	-1.300	-3.400	0.100	-0.100	0.100
Sender Alcohol Use Class 1	0.817	1.333	-	-	-	-	0.983	1.250	-0.017
Sender Alcohol Use Class 2	-	-	-0.416	-0.533	-0.728	-0.836	-	-	-
Edges Class 1	0.219	0.421	0.016	0.055	0.244	0.566	0,055	0.053	0.012
Edges Class 2	0.004	0.026	-	-	0.024	0.071	-	-	-

MJ: Marijuana, Abs Diff: Absolute Difference

Red highlights indicates relative bias in homogeneous parameters above .1, blue highlights indicate relative bias in heterogeneous parameters above .1.

All of the heterogeneous conditions (1 through 4+) demonstrated large relative bias in many of the homogeneous parameters (Highlighted in Table 2.9 in red) and increased raw bias (See Table 2.10). Five effects that had consistent relative bias above .1 in all heterogeneous conditions were Sender Tobacco Use, Sender Anti Social, Receiver Alcohol Use, Absolute Difference in Anti Social and Absolute Difference in Alcohol Use. Absolute Difference in Anti Social and Absolute Difference in Alcohol Use both had true parameters close to 0 (-.08 and -.1 respectively), therefore a certain amount of relative bias is expected. However, when examining the raw biases for the Absolute Difference in Anti Social term a trend emerges. As shown in Table 2.10, there is increased raw bias in the Absolute Difference in Anti Social Behavior parameter as one increases the difference between the latent classes, and this increase in raw bias is greater when one increases the difference between latent classes defined on the edge statistic (Conditions 1 and 1+) than when one increases the difference between latent classes defined on Sender Alcohol Use (Conditions 2 and 2+). When one increases the difference between latent classes defined on both edge parameter and the effect of Sender Alcohol Use, the raw bias in the estimate of the Absolute Difference in Antisocial Behavior is even greater than in any other condition.

Intuitively this is due to the scope of the latent class definition. If a latent class is defined on a structural parameter such as the edge parameter, that effect is present for every node in the network. Therefore the bias caused by specifying a heterogeneous effect of a structural as homogeneous has the potential to spread to any number of other homogeneous parameters. On the other hand, if the latent class is defined only for a subset of nodes, such as when it is defined on a covariate effect, then bias due to mis-specification has a smaller scope, in that only nodes that have the covariate can be influenced by the mis-specification, thus leading to smaller amounts of bias as seen here in conditions 2 and 2+. There is a similar pattern of results for the relative bias in Sender Tobacco Use as well as the Absolute Difference in Anti social behavior.

There is an opposite pattern of bias for homogeneous parameters that involve alcohol use, Receiver Alcohol Use and Absolute Difference in Alcohol Use. For these parameters, the relative biases are greater for Conditions 2 and 2+ then they are for Conditions 1 and 1+.

This is due to the mis-specification of the Sender Alcohol Use parameter forcing the effect into the other Alcohol Use parameters, biasing the estimates. Curiously, the relative bias for either Receiver Alcohol Use or Absolute Difference in Alcohol Use is greater in Condition 2+ (-1.275, -5.100 respectively) than in Condition 3+ (-.650, -3.400 respectively). It appears that the additional mis-specification in the edge parameter is leading to less bias in the alcohol use parameters, yet more bias in general.

Another result of note is the pattern of bias in the GWESP term for Simulations 1 through 3+. With the exception of Condition 2, the relative biases in the estimate of the GWESP term are above .1. This suggests that the GWESP term is sensitive to mis-specification of other heterogeneous effects as homogeneous, and it responds to mis-specification by increasing in magnitude. Intuitively, this is due to transitivity effects being intrinsically confounded with covariate effects. An edge might form because of a similarity on behavior, or due to the influence of shared partners, and if the effect of similarity on behavior is underestimated, the effect of shared partners might be overestimated to compensate. This bias in a structural effect is not replicated in the mutuality parameter, which suggests that mutuality is robust to mis-specifications in the rest of the model.

Additionally, Conditions 4 and 4+ show patterns of bias similar to Conditions 1 and 1+, with relative bias being slightly higher in magnitude for most terms in Conditions 4 and 4+. This suggests that bias due to structural parameters such as GWESP and Edges have a tendency to express itself similarly.

As expected, there was bias in the parameter estimates that were truly heterogeneous (Highlighted in blue in Table 2.9), as it would be impossible to correctly recover the true parameter estimates that were heterogeneous with a single parameter that is homogeneous. Note no relative bias was calculated for the Sender Alcohol Use of Class 1 for Conditions 2 through 3+ as the true value of the parameter was 0.

As for Condition 5, while there was some cases of relative bias being higher than .1

(Sender and Receiver Tobacco Use, and Absolute Difference in Alcohol Use) the rest of the parameters were recovered well. This was expected, as for Condition 5 a homogeneous model is the true model.

Table 2.10: Raw Bias for Homogeneous Models

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1) Class 1	0.127	0.320	0.023	0.099	0.172	0.652	-0.465	-0.614	0.036
GWESP (.1) Class 2	0.127	0.320	0.023	0.099	0.172	0.652	0.335	0.386	0.036
Mutual	0.008	0.019	-0.001	-0.009	0.008	-0.008	0.004	0.003	0.004
Gender Match	-0.015	-0.040	-0.003	-0.009	-0.021	-0.080	-0.038	-0.046	-0.001
Ethnicity Match	-0.055	-0.130	0.004	-0.034	-0.073	-0.221	-0.124	-0.148	0.012
Sender Tobacco Use	0.051	0.079	0.027	0.069	0.079	0.141	0.050	0.063	-0.013
Receiver Tobacco Use	0.013	0.023	-0.001	0.010	0.020	0.048	0.016	0.022	-0.018
Sender Anti Social	0.293	0.501	0.101	0.232	0.366	0.641	0.440	0.527	0.008
Receiver Anti Social	-0.026	-0.042	-0.010	-0.031	-0.021	-0.088	-0.033	-0.045	-0.004
Sender MJ	-0.096	-0.173	-0.031	-0.028	-0.137	-0.187	-0.142	-0.175	-0.003
Receiver MJ	-0.007	-0.009	0.004	-0.002	-0.012	-0.009	-0.001	-0.003	-0.001
Receiver Alcohol Use	-0.004	-0.018	-0.010	-0.051	-0.014	-0.026	-0.015	-0.021	-0.001
Abs Diff Tobacco Use	-0.008	-0.004	0.002	-0.002	-0.009	-0.005	0.010	0.008	0.016
Abs Diff MJ	0.007	0.025	-0.005	0.001	0.012	0.042	0.018	0.025	0.002
Abs Diff Anti Social	0.122	0.191	0.051	0.153	0.169	0.295	0.165	0.193	0.002
Abs Diff Alcohol Use	-0.001	0.002	0.008	0.051	0.013	0.034	-0.001	0.001	-0.001
Sender Alcohol Use Class 1	0.049	0.080	-0.146	-0.467	-0.068	-0.164	0.059	0.075	-0.001
Sender Alcohol Use Class 2	0.049	0.080	0.104	0.533	0.182	0.836	0.059	0.075	-0.001
Edges Class 1	-0.767	-1.368	-0.063	-0.211	-0.854	-1.838	-0.202	-0.193	-0.045
Edges Class 2	-0.017	-0.118	-0.063	-0.211	-0.104	-0.338	-0.202	-0.193	-0.045

MJ: Marijuana, Abs Diff: Absolute Difference

In summary, Tables 2.9 and 2.10 suggest that fitting homogeneous models to networks that are heterogeneous results in unpredictable patterns of high bias in both parameters that are supposed to be heterogeneous, as well as parameters that are indeed homogeneous.

Table 2.11: Mean Standard Errors averaged across 500 trials per condition | Standard Deviations of the Estimated Parameters calculated across 500 trials per condition for Homogeneous Models. Results indicate that there is close agreement between the estimated errors and the simulation errors.

Condition	1	1+	2	2+	3	3+	4	4+	5
GWESP (.1)	0.156|0.160	0.153|0.156	0.172|0.160	0.151|0.154	0.141|0.146	0.144|0.171	0.075|0.081	0.088|0.089	0.204|0.197
Mutual	0.069|0.063	0.071|0.063	0.067|0.060	0.070|0.067	0.070|0.069	0.077|0.069	0.074|0.065	0.072|0.063	0.066|0.060
Gender Match	0.037|0.033	0.038|0.032	0.037|0.033	0.037|0.035	0.038|0.035	0.039|0.032	0.039|0.035	0.038|0.034	0.036|0.032
Ethnicity Match	0.038|0.034	0.038|0.034	0.037|0.033	0.038|0.034	0.039|0.034	0.040|0.034	0.040|0.035	0.038|0.033	0.036|0.032
Sender Tobacco Use	0.071|0.062	0.067|0.054	0.085|0.079	0.088|0.083	0.076|0.073	0.071|0.056	0.076|0.069	0.069|0.066	0.080|0.075
Receiver Tobacco Use	0.069|0.061	0.064|0.053	0.084|0.079	0.087|0.082	0.075|0.070	0.068|0.055	0.075|0.068	0.067|0.059	0.079|0.074
Sender Anti Social	0.128|0.116	0.129|0.121	0.127|0.122	0.130|0.122	0.129|0.116	0.134|0.120	0.131|0.132	0.128|0.118	0.124|0.116
Receiver Anti Social	0.126|0.114	0.127|0.116	0.125|0.117	0.128|0.125	0.128|0.126	0.133|0.117	0.129|0.120	0.126|0.110	0.123|0.109
Sender MJ	0.073|0.060	0.073|0.063	0.075|0.070	0.091|0.082	0.077|0.071	0.091|0.069	0.073|0.066	0.073|0.065	0.072|0.065
Receiver MJ	0.072|0.062	0.072|0.062	0.075|0.068	0.093|0.088	0.077|0.071	0.093|0.074	0.074|0.063	0.073|0.064	0.072|0.065
Receiver Alcohol Use	0.041|0.036	0.041|0.036	0.045|0.041	0.055|0.051	0.045|0.043	0.050|0.043	0.043|0.039	0.042|0.037	0.040|0.037
Abs Diff Tobacco Use	0.069|0.061	0.064|0.052	0.084|0.080	0.088|0.084	0.075|0.071	0.069|0.055	0.074|0.068	0.066|0.062	0.078|0.073
Abs Diff MJ	0.066|0.054	0.067|0.056	0.069|0.061	0.089|0.081	0.072|0.065	0.089|0.068	0.068|0.059	0.067|0.058	0.066|0.058
Abs Diff Anti Social	0.110|0.100	0.110|0.098	0.111|0.105	0.115|0.105	0.112|0.097	0.117|0.096	0.114|0.095	0.111|0.098	0.107|0.096
Abs Diff Alcohol Use	0.034|0.030	0.034|0.030	0.039|0.034	0.050|0.049	0.039|0.034	0.046|0.039	0.036|0.033	0.035|0.031	0.034|0.031
Sender Alcohol Use	0.041|0.036	0.041|0.039	0.044|0.043	0.054|0.053	0.045|0.041	0.050|0.043	0.043|0.040	0.041|0.040	0.040|0.036
Edges	0.192|0.199	0.188|0.195	0.208|0.194	0.185|0.192	0.176|0.179	0.177|0.204	0.110|0.115	0.122|0.124	0.241|0.230

MJ: Marijuana, Abs Diff: Absolute Difference

In terms of standard errors and standard deviation of the simulations for these homogeneous models (Table 2.11), there was close agreement for all terms.

3 EMPIRICAL STUDY

The above simulations were based on an empirical data set of high middle school students. In this empirical example, we examine potential heterogeneity in the effect of alcohol use on peer assortment. Peer assortment on the basis of substance use and other deviant behaviors has been a continually studied topic in clinical psychology (Brechwald & Prinstein, 2011), however, whenever network models are applied to network data to study peer assortment on substance use, researchers have used homogeneous models. Heterogeneous effects of alcohol use on peer assortment could potentially have important implications for network intervention design, as an intervention model based on homogeneous results may not be applicable to all at risk individuals in the intervention population.

We use alcohol use here due to its fairly high use rate among this sample (26.67% of the sample have used alcohol). In future studies of heterogeneous peer assortment and selection effects of substance use, other substances such as marijuana and tobacco should be considered.

Three parameters were used to model each of the effects of alcohol use, marijuana use, tobacco use and anti-social behavior. Indegree parameters measure the effect of the variable on an individuals tendency to receive friendship nominations, outdegree parameters measure the effect of a variable on an individuals tendency to send friendship nominations. Finally, absolute difference terms measure the effect of different values of the variable on the tendency for two individuals to form a friendship.

In the heterogeneous model, we propose a sender-type heterogeneity for the outdegree effect of alcohol use and the absolute difference effect of alcohol use. This can be interpreted as allowing the effect of alcohol use to differ in its effect on an individuals tendency to nominate friends both generally, and with regards to the potential friends alcohol use.

3.1 Data Characteristics and Measures

This empirical data-set consisted of a 151 students who attended the same high school. Students were assessed on a variety of demographics and risk behaviors. The current analysis focuses on substance use of three types, alcohol, tobacco and marijuana, as well as general anti-social behavior. Descriptive statistics of the data are provided in Table 3.1.

Table 3.1: Demographic and Descriptive Statistics

	Male	Female	Missing			
Gender	46%	51.30%	2.60%			
	African-American	Asian	Latino/Hispanic	White	Mixed	Missing
Ethnicity	23.03%	1.32%	18.42%	48.03%	6.58%	2.63%
	Mean	25% Quantile	Median	75% Quantile	% Above 1	Missing
Age	15.07	14.65	15.01	15.33	-	0.66%
Alcohol Use	0.4476	0	0	1	26.97%	5.92%
Tobacco Use	0.1724	0	0	0	6.58%	4.61%
Marijuana Use	0.1781	0	0	0	9.21%	3.95%
Anti social behavior	0.305	0.129	0.258	0.4516	-	4.61%

3.1.1 Measures

Friendship Nominations. Individuals were provided a full roster of the school and were allowed to nominate any number of individuals as friends. These nominations were binary (Yes a friend, or No not a friend), and directed.

Alcohol Use. Alcohol use was assessed with a single item which asked "In the last 6 months on how many days did you have at least one drink of alcohol?" With the following answer choices: 1) 0 days, 2) 1-2 days, 3) 3-5 days, 4) 6-9 days and 5) 10 or more days.

Tobacco Use. Tobacco use was assessed with a single item which asked "In the last 6 months on how many cigarettes did you smoke per day?" With the following answer choices: 1) 0, 2) 1, 3) 2-3 , 4) 6-10 and 5) 10-20, 6) more than a pack.

Marijuana Use. Marijuana use was assessed with a single item which asked "In the last 6 months how many times did you use marijuana?" With the following answer choices: 1) 0, 2) 1-2 times, 3) 3-9 times, 4) 10-19 and 5) 20 or more times.

Anti Social Behavior. Antisocial behavior was assessed with a subset of 31 items from the Youth Self Report scale (Achenbach, 1991). These items assessed behaviors over the previous 6 month period and had text such as "I get in many fights", "I set fires," and "I tease others a lot." Responses were on a 3 point scale with 0 being "Not True", 1 being "Somewhat True", and 2 being "Very/Often True." These items were mean scored for a final Anti-Social behavior Composite.

3.1.2 Missing Data

Due to ERGM modeling inability to handle missing data at the nodal attribute level, multiple imputation was used to account for missingness (Rubin, 1987). 500 covariate datasets were simulated using information on gender, age, ethnicity, substance use and all anti social behavior items. Results for the 500 covariate datasets were combined according to Rubin (1987). There are two competing interests to the use of multiple imputation here. As was mentioned, ERG modeling cannot account for missingness among the nodal covariates, as such one option would be to remove those individuals from the network. However, as networks are highly interdependent systems by definition, this would substantially change the network structure, possibly leading to different results. Multiple imputation however, does not take into a account the dependency between individuals on their nodal attributes, rather assuming them to be independent. This assumption leads to an attenuation of any homophily effects within the network, when nodal covariates are modeled using multiple imputation. In the case of this application, it was decided that the damage to the network structure outweighed the attenuation of homophily, and multiple imputation was used.

3.1.3 Model Specification

Of substantive interest for these analysis is potential heterogeneity in the sender effects of alcohol use, as well as the Sender Effect of the Absolute Difference in Alcohol Use. Additionally, a general heterogeneity was modeled using a heterogeneous edge parameter. This model nearly identical to the model presented in Condition 3 of the simulations was fit. The sole addition to the model was allowing the effect of Absolute Difference in Alcohol Use to vary between the two latent classes. This heterogeneous effect was not included in the simulation conditions due to a lack of heterogeneity in the effect exhibited in the empirical data, as shown below.

Additionally, a homogeneous model was fit to the empirical data.

3.2 Results of the Empirical Analysis.

Table 3.2: Empirical Results for both heterogeneous model and homogeneous model. Estimates are in mean log-odds metric calculated across 500 multiply imputed datasets with (multiple imputation standard error)

Parameter	Heterogeneous Model	Homogeneous Model
GWESP (alpha = .1)	0.97 (0.07)*	1.25 (0.07)*
Mutuality	2.34 (0.14)*	2.26 (0.15)*
Gender Match	0.52 (0.06)*	0.44 (0.06)*
Ethnicity Match	0.85 (0.06)*	0.47 (0.06)*
Sender Tobacco Use	0.13 (0.11)	0.29 (0.10)*
Receiver Tobacco Use	-0.11 (0.10)	-0.05 (0.09)
Sender Anti Social	-0.64 (0.21)*	-0.23 (0.20)
Receiver Anti Social	0.24 (0.20)	0.24 (0.20)
Sender MJ	0.33 (0.11)*	0.14 (0.10)
Receiver MJ Use	0.21 (0.11)	0.19 (0.10)
Receiver Alcohol Use	0.04 (0.07)	-0.02 (0.06)
Abs Difference Anti Social	-0.08 (0.11)	-0.02 (0.18)
Abs Difference Tobacco Use	-0.21 (0.10)*	-0.13 (0.10)
Abs Difference MJ Use	-0.39 (0.19)*	-0.18 (0.10)
Abs Difference Alcohol Use Class 1	-0.01 (0.08)	-0.01 (0.06)
Abs Difference Alcohol Use Class 2	-0.05 (0.07)	
Sender Alcohol Use Class 1	-0.21 (0.10)*	0.02 (0.06)
Sender Alcohol Use Class 2	0.06 (0.07)	
Edges Class 1	-5.28 (0.11)*	-5.09 (0.10)*
Edges Class 2	-4.22 (0.12)*	

MJ: Marijuana, Abs Difference: Absolute Difference

*: $p < .05$

Table 3.2 contains the results for the heterogeneous and homogeneous analysis. Of note in these results is the effect of Sender Alcohol Use in Class 1 versus the effect of Alcohol Use in Class 2 . These effects are significantly different from one another (z = -2.21, p <.05), suggesting that for members of class 1, alcohol use reduces the number of friendship nominations sent out while for members of Class 2, alcohol use does not effect the number of friendship nominations sent out. Additionally, there is a significant difference between the edges effect of Class 1 and Class 2 (z = 6.51, p <.05). This suggests that members of Class 1 have fewer friendship nominations made in general, while members of Class 2 make more friendship nominations.

There are several significant homogeneous effects in the heterogeneous model. The effect of Sender Anti Social Behavior is negative suggesting that individuals with more antisocial behavior nominate fewer friends. The effects of Absolute Difference in Tobacco and Marijuana Use are both negative suggesting that individuals who are different on their usage of those substances are less likely to nominate each other as friends. Finally there was a positive significant effect of Sender Marijuana Use suggesting that individuals who use more Marijuana make more friendship nominations in general.

There are expected positive significant findings for GWESP, mutuality, gender and ethnicity matching. The significant GWESP finding suggests that individuals who share friends are more likely to be friends themselves. Additionally there is a strong effect for mutuality suggesting that reciprocal friendship nominations are the norm in this network.

As for the differences between the heterogeneous model and the homogeneous model, it is apparent that the modeling of heterogeneous effects of alcohol use and the edge parameter clarifies several of the homogeneous parameters. Specifically, the effects for Sender Anti Social Behavior, Sender Marijuana Use, Absolute Difference in Tobacco Use and Absolute Difference in Marijuana Use are significant in the Heterogeneous Model and not significant in the Homogeneous Model. Interestingly, one effect, that of Sender Tobacco

Use, was significant in the homogeneous model and not significant in the heterogeneous model.

It is worthwhile to note that significantly fewer individuals were classified into class 2 than into class 1, with the average proportion of individuals in class 2 across all the multiply imputed datasets being 27.8%. Class separation on the basis of class membership probability was exceptionally good for every multiply imputed dataset. The average probability of membership in class 1 for individuals with a higher than .5 probability of being in class 1 was .939, while the average probability of class membership in class 2 for individuals with a higher than .5 probability of being in class 2 was .9671. This level of class separation is in part due to the classification EM algorithm used in estimation, which tends to find the latent class labeling that is the most informative rather than the latent class probabilities that maximize the labeling.

In summary, the findings support a level of heterogeneity in the effect of alcohol use, specifically for the effect of alcohol use on general outdegree. Indeed, some individuals are less likely to nominate friends the more alcohol they drink, while other individuals exhibit no effect of alcohol use on friendship formation. Of broader interest, there are more significant findings in the heterogeneous model as compared to the homogeneous model. This suggests that the inclusion of the heterogeneous model parameters clarified the effects of the other parameters.

4 DISCUSSION

This thesis presented the Sender/Receiver Latent Class ERGM, provided a estimation routine, presented simulations that demonstrated its ability to detect latent class structures in networks, and demonstrated the consequences of failing to model latent class structures in network. Finally, this thesis presented an empirical example rooted in adolescent peer networks that demonstrated this models ability to obtain reasonable results from real, rather than simulated data.

There are several broad results that can be gleaned from the simulations. First, recovery of the latent class structures is sensitive to the difference between the latent classes, as demonstrated by the increased difference conditions in the simulations. This is to be expected; however, caution is advised when using these models on empirical data as the models will return solutions regardless of underlying differences between the latent classes. One solution to this problem is examining the *a posteriori* probabilities of class membership given the returned set of parameter estimates. If there is a large amount of uncertainty in the classifications of individuals, then the model should be reassessed. Simulation results suggest that the classification probabilities should be quite high for correctly classified individuals.

There is also caution warranted when analyzing data with homogeneous models. As shown in the simulations, mis-specifying structural parameters such as edges or GWESP can result in wide-ranging estimate bias of many other model parameters, where this bias is generally unpredictable in its direction. Additional caution is warranted when looking at the effects of covariates as a mis-specification of one covariate effect (such as a sender effect) as homogeneous when really there are heterogeneous effects leads to bias in other effects of the same covariate (such as absolute difference). This is particularly concerning as the direction of the bias is unpredictable, and therefore mis-specification could lead to spurious findings.

As another finding of particular interest, mis-specifying a heterogeneous model to a truly homogeneous network did not result in any real bias, other than in the mis-specified terms. Even with the mis-specified terms the differences in effect size were small. This suggests that researchers can use heterogeneous models as an exploratory method, without worry that a heterogeneous mis-specification of a homogeneous network would result in biased estimates.

The empirical example demonstrates the utility of these models for answering substantive questions. Here, that question was: does the effect of alcohol use on the number of friends one nominates differ between individuals? With this sample, there appears to be at least some level of heterogeneity in the effect of alcohol use on the number of friends an individual nominates. Furthermore, by modeling heterogeneity, several effects were clarified that would have not been significant if a homogeneous model was run. These findings further emphasize the importance of modeling individual differences in network data.

There were several limitations to the Sender/Receiver latent class ERGMs and to this thesis as a whole. First, the Sender/Receiver Latent Class ERGM requires a directed network. In many cases, the network, and the relation that defines the edges in the network would not be directed, and these models would not be able to be fit. Secondly, these models require the analyst to make a choice to model either Sender latent classes, or Receiver latent classes, but not both. This restricts the type of heterogeneity that an analyst can model. Finally, as developed the Sender/Receiver Latent Class ERGM is only for cross-sectional data, which limits its use in answering substantive questions.

As for the thesis itself, one key limitation is the choice not to manipulate sample sizes. Adjusting sample sizes in networks is a complex matter, as a network by definition is not a sample of independent observations. However, larger networks imply more information, and it would be expected that with larger networks, latent class recovery would be simpler.

As another limitation, the routine for estimating Sender/Receiver Latent Class ERGMs rely on a classification type EM algorithm. This algorithm is well known to perform badly when classes are ill-separated, or have radically different proportions of cases (Bryant, 1991; Celeux & Govaert, 1993). Well defined classes are not often the case in psychological research, as such the use of the classification likelihood method as outlined in this thesis would not be warranted in cases where the class separation is not *a priori* thought to be high. Additional simulations should be done comparing the performance of mixture and classification type EM algorithms at estimating these models.

However, these limitations open up several new directions for future work. First, and possibly most importantly, would be to expand the estimation to use mixture modeling as opposed to classification likelihood. This would allow researchers to investigate scenarios where class separation is small. The next step would be extending these models to fit general latent class ERGMs, rather than restricting the latent class definition to be sender or receiver. This likely would require several approximations, and therefore a Bayesian approach might be warranted. A general Latent Class ERGM would very much resemble the ERGM models proposed by Schweinberger and Handcock (2014), however, would not focus so much on latent community detection (though they would be capable of it). Additionally, a general latent class ERGM would be able to model undirected networks. Another extension would be to develop a longitudinal version. Adapting the separable temporal ERGMs (Krivitsky & Handcock, 2014) which can model peer selection processes over time, would allow analysts to investigate heterogeneous peer selection processes.

Overall, this thesis introduces a new way of modeling network data and a different definition of heterogeneity on networks that could better serve substantive researchers studying a variety of heterogeneous and latent network-rooted behaviors and traits, such as in adolescent peer influence of substance abuse (eg., Brechwald & Prinstein, 2011), in clinical domains such as autism research (eg., Gilman, et al., 2011), and also in social, cognitive, and affective neuroscience (Wager et al., 2015). Further work is needed to refine and expand the latent class ERGM so that psychological and social scientists can advance our understanding of ideographic variation in network settings.

REFERENCES

Achenbach, T. M. (1991). Manual for the Youth Self Report and 1991 Profile. *Burlington VT*.

Ali, M. M., & Dwyer, D. S. (2011). Estimating peer effects in sexual behavior among adolescents. *Journal of Adolescence*, *34*(1), 183–190. doi: 10.1016/j.adolescence.2009.12.008

Bauer, D. J. (2011). Evaluating Individual Differences in Psychological Processes. *Current Directions in Psychological Science*, *20*(2), 115–118. doi: 10.1177/0963721411402670

Bauer, D. J., & Cai, L. (2008). Consequences of Unmodeled Nonlinear Effects in Multilevel Models. *Journal of Educational and Behavioral Statistics*, *34*(1), 97–114. doi: 10.3102/1076998607310504

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society. Series B (...*, *36*(2), 192–225.

Brechwald, W. A., & Prinstein, M. J. (2011). Beyond homophily: A decade of advances in understanding peer influence processes. *Journal of Research on Adolescence*, *21*(1), 166–179. doi: 10.1111/j.1532-7795.2010.00721.x

Bryant, P. G. (1991, January). Large-sample results for optimization-based clustering methods. *Journal of Classification*, *8*(1), 31–44. doi: 10.1007/BF02616246

Celeux, G., & Govaert, G. (1993, October). Comparison of the mixture and the classification maximum likelihood in cluster analysis. *Journal of Statistical Computation and Simulation*, *47*(3-4), 127–146. doi: 10.1080/00949659308811525

Chatterjee, S., & Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Annals of Statistics*. doi: 10.1307/aos/1383661269

Comets, F., & Janžura, M. (1998). A central limit theorem for conditionally centred random fields with an application to Markov fields. *Journal of Applied Probability*, *35*(3), 608–621. doi: 10.1239/jap/1032265209

Daudin, J.-J., Picard, F., & Robin, S. (2006). A mixture model for random graphs. *Stat. Comput.*, *18*(2), 173–183. doi: 10.1007/s11222-007-9046-7

DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, *5*(2), 249–282. doi: 10.1007/BF01897167

Efron, B. (1978). The Geometry of Exponential Families. *The Annals of Statistics*, *6*(2), 362–376. doi: 10.1214/aos/1176344130

Erdos, P., & Renyi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, *81*(395), 832–842. doi: 10.2307/2289017

Gilman, S. R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., & Vitkup, D. (2011). Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses. *Neuron*, *70*(5), 898–907. doi: 10.1016/j.neuron.2011.05.021

Govaert, G., & Nadif, M. (1996). Comparison of the mixture and the classification maximum likelihood in cluster analysis with binary data. *Computational Statistics and Data Analysis*, *23*(1), 65–81. doi: 10.1016/S0167-9473(96)00021-7

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). statnet: Software tools for the Statistical Modeling of Network Data. Retrieved from `http://cran.r-project.org/package=statnethttp://statnetproject.org`

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*(1), 36–53. doi: 10.1037/1082-989X.11.1.36

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098. doi: 10.1198/016214502388618906

Holland, P., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Associationerican Statistical ...*, *76*(373), 33–50.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008, March). Goodness of Fit of Social Network Models. , *103*(481), 248–258. doi: 10.1198/016214507000000446

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). Interaction effects in multiple regression. *Sage University Paper Series on Quantitative Applications in the Social Sciences*(07-072).

Jaccard, J., Wan, C. K., & Turrisi, R. (1990). The Detection and Interpretation of Interaction Effects Between Continuous Variables in Multiple Regression. *Multivariate Behavioral Research*, *25*(4), 467–478. doi: 10.1207/s15327906mbr2504_4

Kearns, M., Mansour, Y., & Ng, A. Y. (1997). An Information-Theoretic Analysis of Hard and Soft Assignment Methods for Clustering. *Proc. of Conference on Uncertainty in*

Artificial Intelligence, 282–293.

Kiuru, N., Burk, W. J., Laursen, B., Salmela-Aro, K., & Nurmi, J. E. (2010). Pressure to drink but not to smoke: Disentangling selection and socialization in adolescent peer networks and peer groups. *Journal of Adolescence*, *33*(6), 801–812. doi: 10.1016/j.adolescence.2010.07.006

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *76*(1), 29–46. doi: 10.1111/rssb.12014

Lubbers, M. J., & Snijders, T. A. B. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social Networks*, *29*(4), 489–507. doi: 10.1016/j.socnet.2007.03.002

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proc. Symposium on Mathematical Statistics and Probability*, 281–297. doi: citeulike-article-id:6083430

McLachlan, G., & Peel, D. (2005, January). ML Fitting of Mixture Models. In *Finite mixture models (wiley series in probability and statistics)* (pp. 40–80). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi: 10.1002/0471721182.ch2

Mercken, L., Steglich, C., Sinclair, P., Holliday, J., & Moore, L. A. R. (2012). A longitudinal social network analysis of peer influence, peer selection, and smoking behavior among adolescents in British schools. *Health Psychology*, *31*(4). doi: 10.1037/a0026876

Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218. doi: 10.1207/s15366359mea0204_1

Nowicki, K., & Snijders, T. A. B. (2001). *Estimation and Prediction for Stochastic Blockstructures* (Vol. 96) (No. 455). doi: 10.1198/016214501753208735

Osgood, D. W., Feinberg, M. E., & Ragan, D. T. (2015). Social Networks and the Diffusion of Adolescent Problem Behavior: Reliable Estimates of Selection and Influence from Sixth Through Ninth Grades. *Prevention Science*. doi: 10.1007/s11121-015-0558-7

Osgood, D. W., Ragan, D. T., Wallace, L., Gest, S. D., Feinberg, M. E., & Moody, J. (2013). Peers and the emergence of alcohol use: Influence and selection processes in adolescent friendship networks. *Journal of Research on Adolescence*, *23*(3), 500–512. doi: 10.1111/jora.12059

Pu, W., Choi, J., Hwang, Y., & Amir, E. (2015). A Deterministic Partition Function

Approximation for Exponential Random Graph Models. *International Joint Conference on Artificial Intelligence*.

Robins, G., Elliott, P., & Pattison, P. (2001, January). Network models for social selection processes. *Social Networks*, *23*(1), 1–30. doi: 10.1016/S0378-8733(01)00029-6

Rubin, A. P. D., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. B*, *39*(1), 1–38.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. *Harvard University*(JOHN WILEY & SONS), 15–19. doi: 10.1002/9780470316696

Schaefer, D. R., Simpkins, S. D., Vest, A. E., & Price, C. D. (2011). The contribution of extracurricular activities to adolescent friendships: new insights through social network analysis. *Developmental psychology*, *47*(4), 1141–1152. doi: 10.1037/a0024091

Schweinberger, M., & Handcock, M. S. (2015). Local dependence in random graph models: characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *77*(3), 647–676. doi: 10.1111/rssb.12081

Shin, H.-S., Valente, T. W., Riggs, N. R., Huh, J., Spruijt-Metz, D., Chou, C.-P., & Ann Pentz, M. (2014). The interaction of social networks and child obesity prevention program effects: the pathways trial. *Obesity*, *22*(6), 1520–6. doi: 10.1002/oby.20731

Snijders, T. A., van de Bunt, G. G., & Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, *32*(1), 44–60. doi: 10.1016/j.socnet.2009.02.004

Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological methods*, *9*(3), 386–396. doi: 10.1037/1082-989X.9.3.386

Symons, M. J. (1981, March). Clustering Criteria and Multivariate Normal Mixtures. *Biometrics*, *37*(1), 35. doi: 10.2307/2530520

van Duijn, M. A. J., Gile, K. J., & Handcock, M. S. (2009). A Framework for the Comparison of Maximum Pseudo Likelihood and Maximum Likelihood Estimation of Exponential Family Random Graph Models. *Social Networks*, *31*(1), 52–62. doi: 10.1016/j.socnet.2008.10.003

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian Model of Category-Specific Emotional Brain Responses. *PLOS Computational Biology*, *11*(4), e1004066. doi: 10.1371/journal.pcbi.1004066

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. *Psychometrika*, *61*(3), 401–425.

doi: 10.1007/BF02294547

Wedel, M., & DeSarbo, W. S. (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification*, *12*(1), 21–55. doi: 10.1007/BF01202266